(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 03/106648 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/18934

(22) International Filing Date: 16 June 2003 (16.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/389,327 14 June 2002 (14.06.2002) US

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, San Francisco, CA 94080 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **SALCEDA, Susana** [AR/US]; 4118 Cresendo Avenue, San Jose, CA 95136 (US). **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **TURNER, Leah, R.** [US/US]; 939 Rosette Court, Sunnyvale, CA 94086 (US). **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) **Agents: LICATA, Jane, Massey** et al.; Licata & Tyrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/106648 A2

(54) **Title:** COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) **Abstract:** The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions containing the nucleic acid molecules, polypeptides, antibodies, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic breast tissue, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast, identifying breast tissue, monitoring and modifying breast tissue development and differentiation, and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

BACKGROUND OF THE INVENTION

Excluding skin cancer, breast cancer, also called mammary tumor, is the most common cancer among women, accounting for a third of the cancers diagnosed in the United States. One in nine women will develop breast cancer in her lifetime and about 192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths. Bevers, *Primary Prevention of Breast Cancer*, in BREAST CANCER, 20-54 (Kelly K Hunt et al., ed., 2001); Kochanek et al., Nat'l.Vital Statistics Reports 49(1):14 (2001).

In the treatment of breast cancer, there is considerable emphasis on detection and risk assessment because early and accurate staging of breast cancer has a significant impact on survival. For example, breast cancer detected at an early stage (stage T0, discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not detected until a late stage (i.e., stage T4), the five-year survival rate is reduced to 13%. AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998). Some detection techniques, such as mammography and biopsy, involve increased

discomfort, expense, and/or radiation, and are prescribed only to patients with an increased risk of breast cancer.

Current methods for predicting or detecting risk of breast cancer are not optimal. One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regiments for high risk patients. A patient's risk of breast cancer has been positively associated with increasing age, nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer limits their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than increasing age. NIH Publication No. 00-1556 (2000).

Current screening methods for detecting cancer, such as self-examination, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent their widespread adoption. Self-examination, while useful, is unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by palpitation. Ultrasound measurements require skilled operators at an increased expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation used in mammography because prior chest radiation is a factor associated with an increased incidence of breast cancer.

At this time, there are no adequate methods of breast cancer prevention. The current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy before cancer diagnosis), which are drastic measures that limit their adoption even among women with, increased risk of breast cancer. Bevers, *supra*.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., JAMA 249:1881 (1983)) MUC-1 (Frische and Liu, J. Clin. Ligand 22:320 (2000)), HER-2/neu (Haris et al., Proc.Am.Soc.Clin.Oncology. 15:A96 (1996)), uPA, PAI-1, LPA, LPC, RAK

and BRCA (Esteva and Fritsche, *Serum and Tissue Markers for Breast Cancer*, in BREAST CANCER, 286-308 (2001)).

Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. There are a variety of different schemes for staging breast cancer. One is known as the TNM staging system in which T stands for tumor size, N stands for node involvement and M stands for metastasis. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ, intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 is characterized by tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T3. A T4 stage tumor may be any size with an extension to either the chest wall or the skin. Within stage T4, T4a indicates extension of the tumor to the chest wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. AJCC Cancer Staging Handbook pp. 159-70 (Irvin D. Fleming et al. eds., 5th ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., Breast Cancer Research and Treatment 7:147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., J.Nat'l.Cancer Inst. 90:1346 (1998); Paik et al., J.Nat'l.Cancer Inst. 90:1361 (1998); Hutchins et al., Proc.Am.Soc.Clin.Oncology 17:A2 (1998).; and Simpson et al., J.Clin.Oncology 18:2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or to

other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., J. Clin. Oncology 18:2059 (2000). Generally, pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids that could differentiate between different stages of invasion. Progress in this field will allow more rapid and reliable methods for treating breast cancer patients.

Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., Annals of Internal Medicine 127:1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for early stage breast cancer. Fisher et al., J. of Clinical Oncology 16:441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may be employed. These treatment methods are equally effective in the early stages of breast cancer.

Another staging scheme is Stage I, II, III and IV. In this scheme, Stage I is characterized as having a tumor of 2 cm or less and no lymph node involvement or metastasis. Stage II is characterized by a tumor of 2 cm to 5 cm and local or no lymph node involvement and no metastasis. Stage III is greater than 5 cm and local lymph node involvement and no metastasis. Stage IV is a metastatic tumor with no regard for size or lymph node involvement. Patients with Stage I and Stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and Stage IV patients. Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent

restaging because cancer is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998).

To provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence that can be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page et al., Cancer 75:1219 (1995); Fisher et al., Cancer 75:1223 (1995); Silverstein et al., Cancer 77:2267 (1996).

As discussed above, each of the methods for diagnosing and staging breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of breast cancers to optimize treatment methods. Finally, there is a need for sensitive molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of breast cancers following remission.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

SUMMARY OF THE INVENTION

The present invention solves many needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, agonists and antagonists that may be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast; identify and monitor breast tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered breast tissue for treatment and research.

One aspect of the present invention relates to nucleic acid molecules that are specific to breast cells, breast tissue and/or the breast organ. These breast specific nucleic acids (BSNAs) may be a naturally occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. If the BSNA is genomic DNA, then the BSNA is a breast specific gene (BSG). If the BSNA is RNA, then it is a breast specific transcript encoded by a BSG. Due to alternative splicing and transcriptional modification one BSG may encode for multiple breast specific RNAs. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast. More preferred is a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 95-156. In another preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-94. For the BSNA sequences listed herein, DEX0432_001.nt.1 corresponds to SEQ ID NO: 1. For sequences with multiple splice variants, the parent sequence DEX0432_001.nt.1, will be followed by DEX0432_001.nt.2, etc. for each splice variant. The sequences off the corresponding peptides are listed as DEX0432_001.aa.1, etc. For the mapping of all of the nucleotides and peptides, see the table in the Example 1 section below.

This aspect of the present invention also relates to nucleic acid molecules that selectively hybridize or exhibit substantial sequence similarity to nucleic acid molecules encoding a breast Specific Protein (BSP), or that selectively hybridize or exhibit substantial sequence similarity to a BSNA. In one embodiment of the present invention the nucleic acid molecule comprises an allelic variant of a nucleic acid molecule encoding a BSP, or an allelic variant of a BSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid sequence that encodes a BSP or a part of a nucleic acid sequence of a BSNA.

In addition, this aspect of the present invention relates to a nucleic acid molecule further comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a BSNA or the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a BSP.

Another aspect of the present invention relates to vectors and/or host cells comprising a nucleic acid molecule of this invention. In a preferred embodiment, the nucleic acid molecule of the vector and/or host cell encodes all or a fragment of a BSP. In another preferred embodiment, the nucleic acid molecule of the vector and/or host cell

comprises all or a part of a BSNA. Vectors and host cells of the present invention are useful in the recombinant production of polypeptides, particularly BSPs of the present invention.

Another aspect of the present invention relates to polypeptides encoded by a nucleic acid molecule of this invention. The polypeptide may comprise either a fragment or a full-length protein. In a preferred embodiment, the polypeptide is a BSP. However, this aspect of the present invention also relates to mutant proteins (muteins) of BSPs, fusion proteins of which a portion is a BSP, and proteins and polypeptides encoded by allelic variants of a BSNA as provided herein.

A further aspect of the present invention is a novel splice variant which encodes an amino acid sequence that provides a novel region to be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer. The novel amino acid sequence may lead to a unique protein structure, protein subcellular localization, biochemical processing or function. This information can be used to directly or indirectly facilitate the generation of additional or novel therapeutics or diagnostics. The nucleotide sequence in this novel splice variant can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

Another aspect of the present invention relates to antibodies and other binders that specifically bind to a polypeptide of the instant invention. Accordingly antibodies or binders of the present invention specifically bind to BSPs, muteins, fusion proteins, and/or homologous proteins or polypeptides encoded by allelic variants of an BSNA as provided herein.

Another aspect of the present invention relates to agonists and antagonists of the nucleic acid molecules and polypeptides of this invention. The agonists and antagonists of the instant invention may be used to treat breast cancer and non-cancerous disease states in breast and to produce engineered breast tissue.

Another aspect of the present invention relates to methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. Such methods are useful in identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast. Such methods are also useful in identifying and/or monitoring breast tissue. In addition, measurement of levels of one or more of the nucleic acid molecules of this invention may be useful for diagnostics as part

of panel in combination with known other markers, particularly those described in the breast cancer background section above.

Another aspect of the present invention relates to use of the nucleic acid molecules of this invention in gene therapy, for producing transgenic animals and cells, and for producing engineered breast tissue for treatment and research.

Another aspect of the present invention relates to methods for detecting polypeptides this invention, preferably using antibodies thereto. Such methods are useful to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast. In addition, measurement of levels of one or more of the polypeptides of this invention may be useful to identify, diagnose, monitor, stage, image breast cancer in combination with known other markers, particularly those described in the breast cancer background section above. The polypeptides of the present invention can also be used to identify and/or monitor breast tissue, and to produce engineered breast tissue.

Yet another aspect of the present invention relates to a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences. In addition, the computer records regarding the nucleic acid and/or amino acid sequences and/or measurements of their levels may be used alone or in combination with other markers to diagnose breast related diseases.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various

general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4^{th} Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999).

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages.

Nucleotides are represented by single letter symbols in nucleic acid molecule sequences. The following table lists symbols identifying nucleotides or groups of nucleotides that may occupy the symbol position on a nucleic acid molecule. *See* Nomenclature Committee of the International Union of Biochemistry (NC-IUB),

Nomenclature for incompletely specified bases in nucleic acid sequences, Recommendations 1984., *Eur J Biochem.* 150(1):1-5 (1985).

Symbol	Meaning	Group/Origin of Designation	Complementary Symbol
a	a	Adenine	t/u
g	g	Guanine	c
c	c	Cytosine	g
t	t	Thymine	a
u	u	Uracil	a
r	g or a	puRine	y
y	t/u or c	pYrimidine	r
m	a or c	aMino	k
k	g or t/u	Keto	m
s	g or c	Strong interactions 3H-bonds	w
w	a or t/u	Weak interactions 2H-bonds	s
b	g or c or t/u	not a	v
d	a or g or t/u	not c	h
h	a or c or t/u	not g	d
v	a or g or c	not t, not u	b
n	a or g or c or t/u, unknown, or other	any	n

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to

occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such

oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally occurring nucleotide" referred to herein includes naturally occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), and U.S. Patent No. 5,151,510, the disclosure of which is hereby incorporated by reference in its entirety.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences that are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art that can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998)). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, double stranded RNA (dsRNA) inhibition (RNAi), combination of triplex and antisense, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least

about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists between a first and second nucleic acid sequence when the first nucleic acid sequence or fragment thereof hybridizes to an antisense strand of the second nucleic acid, under selective hybridization conditions. Typically, selective hybridization will occur between the first nucleic acid sequence and an antisense strand of the second nucleic acid sequence when there is at least about 55% sequence identity between the first and second nucleic acid sequences— preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% — over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) –
0.63 (% formamide) - (600/l) where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) +
11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) +

11.8 (fraction $G + C)^2$ - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

$T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction } G+C) -(600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well known in the art. *See*, *e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and are specified by commercial

suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra.*

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies. In another aspect, the invention is directed to single exon probes based on the BSNAs disclosed herein.

In one embodiment, the term "microarray" refers to a "nucleic acid microarray" having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the devices so called in Schena (ed.), <u>DNA Microarrays: A Practical Approach (Practical Approach Series</u>), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), <u>Microarray Biochip: Tools and Technology</u>, Eaton Publishing Company/BioTechniques Books Division (2000). Additionally, these nucleic acid

microarrays include substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Patent Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, 5,405,783, the disclosures of which are incorporated herein by reference in their entireties.

In an alternative embodiment, a "microarray" may also refer to a "peptide microarray" or "protein microarray" having a substrate-bound collection of plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray ,may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, aptamers, which can specifically detect the binding of the polypeptides of this invention. The array may be based on autoantibody detection to the polypeptides of this invention, see Robinson *et al., Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, WO 97/42507 and U.S. Patent Nos. 6,268,210, 5,766,960, 5,143,854, the disclosures of which are incorporated herein by reference in their entireties.

In addition, determination of the levels of the BSNA or BSP may be made in a multiplex manner using techniques described in WO 02/29109, WO 02/24959, WO 01/83502, WO01/73113, WO 01/59432, WO 01/57269, WO 99/67641, the disclosures of which are incorporated herein by reference in their entireties.

The term "mutant", "mutated", or "mutation" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci

within a nucleic acid sequence. In a preferred embodiment of the present invention, the nucleic acid sequence is the wild type nucleic acid sequence encoding a BSP or is a BSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process that enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process that involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids that lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993).

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is either contiguous with the gene of interest to control the gene of interest, or acts in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences that are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences that control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial

artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refers to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence is meant to be inclusive of all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally occurring and non-naturally occurring proteins and polypeptides, as well as polypeptide fragments and polypeptide mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a BSP encoded by a nucleic acid molecule of the instant invention, or a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be determined by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art for purification.

The term "fragment" when used herein with respect to polypeptides of the present invention refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length BSP. In a preferred embodiment, the fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally occurring polypeptide. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40

or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" when used herein with respect to polypeptides of the present invention refers to a polypeptide which is substantially similar in primary structural sequence to a BSP but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the BSP. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, ^{14}C and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*.

The term "fusion protein" refers to polypeptides of the present invention coupled to a heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence that encodes the polypeptide or a fragment thereof in frame with a nucleic

acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --CH(OH)CH_2--, and –CH_2SO--, by methods well known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992)). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

The term "mutant" or "mutein" when referring to a polypeptide of the present invention relates to an amino acid sequence containing substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a BSP. A

mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to a BSP. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as GAP or BESTFIT or other variation Smith-Waterman alignment. *See*, T. F. Smith and M. S. Waterman, J. Mol. Biol. 147:195-197 (1981) and W.R. Pearson, Genomics 11:635-650 (1991).

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991).

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991). Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

By "homology" or "homologous" when referring to a polypeptide of the present invention it is meant polypeptides from different organisms with a similar sequence to the encoded amino acid sequence of a BSP and a similar biological activity or function. Although two polypeptides are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the polypeptides. Instead, the term "homologous" is defined to mean that the two polypeptides have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous polypeptide is one that exhibits 50% sequence similarity to BSP, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous polypeptides that exhibit 80%, 85% or 90% sequence similarity to a BSP. In a yet more preferred embodiment, a homologous polypeptide exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative

substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994).

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992). A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997). Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)

Cost to extend a gap: 1 (default
Max. alignments: 100 (default)
Word size: 11 (default)
No. of descriptions: 100 (default)
Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Algorithms other than blastp for database searching using amino acid sequences are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. A Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; a $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; a Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.,* Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" as used herein it is meant the ability of the antibody to bind to a first molecular species in preference to binding to other

molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which VL and VH regions are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.,* Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that

normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "breast specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the breast as compared to other tissues in the body. In a preferred embodiment, a "breast specific" nucleic acid molecule or polypeptide is detected at a level that is 1.5-fold higher than any other tissue in the body. In a more preferred embodiment, the "breast specific" nucleic acid molecule or polypeptide is detected at a level that is 2-fold higher than any other tissue in the body, more preferably 5-fold higher, still more preferably at least 10-fold, 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the breast or to breast cells or tissue or that are derived from such nucleic acid molecules. These isolated breast specific nucleic acids (BSNAs) may comprise cDNA genomic DNA, RNA, or a combination thereof, a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. A BSNA may be derived from

an animal. In a preferred embodiment, the BSNA is derived from a human or other mammal. In a more preferred embodiment, the BSNA is derived from a human or other primate. In an even more preferred embodiment, the BSNA is derived from a human.

In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast, a breast-specific polypeptide (BSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 95-156. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-94. Nucleotide sequences of the instantly described nucleic acid molecules were determined by assembling several DNA molecules from either public or proprietary databases. Some of the underlying DNA sequences are the result, directly or indirectly, of at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA).

Nucleic acid molecules of the present invention may also comprise sequences that selectively hybridizes to a nucleic acid molecule encoding a BSNA or a complement or antisense thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may or may not encode a BSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a BSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 95-156. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1-94 or the antisense sequence thereof. Preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a BSP under low stringency conditions. More preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a BSP under moderate stringency conditions. Most preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a BSP under high stringency conditions. In a preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic

acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 95-156. In a more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1-94.

Nucleic acid molecules of the present invention may also comprise nucleic acid sequences that exhibit substantial sequence similarity to a nucleic acid encoding a BSP or a complement of the encoding nucleic acid molecule. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule encoding human BSP. More preferred is a nucleic acid molecule exhibiting substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 95-156. By substantial sequence similarity it is meant a nucleic acid molecule having at least 60% sequence identity with a nucleic acid molecule encoding a BSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 95-156, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a BSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred in this embodiment is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a BSP.

The nucleic acid molecules of the present invention are also inclusive of those exhibiting substantial sequence similarity to a BSNA or its complement. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO: 1-94. By substantial sequence similarity it is meant a nucleic acid molecule that has at least 60% sequence identity with a BSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1-94, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. More preferred is a nucleic acid molecule that has at least 90% sequence identity with a BSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a BSNA.

Nucleic acid molecules that exhibit substantial sequence similarity are inclusive of sequences that exhibit sequence identity over their entire length to a BSNA or to a nucleic acid molecule encoding a BSP, as well as sequences that are similar over only a part of its length. In this case, the part is at least 50 nucleotides of the BSNA or the nucleic acid molecule encoding a BSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 95-156 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1-94. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule from a human, when the BSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a BSNA. In a preferred embodiment, the substantially similar nucleic acid molecule is an BSNA.

The nucleic acid molecules of the present invention are also inclusive of allelic variants of a BSNA or a nucleic acid encoding a BSP. For example, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes and the sequence determined from one individual of a species may differ from other allelic forms present within the population. More than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – Variants with small deletions and insertions of more than a single nucleotide are

also found in the general population, and often do not alter the function of the protein. In addition, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a BSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a BSP comprising an amino acid sequence of SEQ ID NO: 95-156. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a BSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1-94. Also preferred is that the allelic variant is a naturally occurring allelic variant in the species of interest, particularly human.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences comprising a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a BSP. In a preferred embodiment, the part encodes a BSP. In one embodiment, the nucleic acid molecule comprises a part of a BSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a BSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that is an allelic variant of a BSNA. In yet another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that encodes a BSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences that encode fusion proteins, homologous proteins, polypeptide fragments, muteins and polypeptide analogs, as described *infra*.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences containing modifications of the native nucleic acid molecule. Examples of such modifications include, but are not limited to, nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that may be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is

used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

Accordingly, in one embodiment, a nucleic acid molecule may include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. The labeled nucleic acid molecules are particularly useful as hybridization probes.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dGTP, α-^{32}P-dTTP, α-^{32}P-3'dATP, α-^{32}P-ATP, α-^{32}P-CTP, α-^{32}P-GTP, α-^{32}P-UTP, α-^{35}S-dATP, γ-^{35}S-GTP, γ-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Biosciences, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000).

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp.,

Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules of the present invention can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and Peptide Nucleic Acids (PNA) to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994). Alternatively, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.,* Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*,

Nucleic Acids Symp. Ser. (37): 255-6 (1997); and U.S. Patent Nos. 5,846,726, 5,925,517, 5,925,517, 5,723,591 and 5,538,848, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the present invention may also be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents – Symposium No. 209, John Wiley & Son Ltd (1997). Such altered internucleoside bonds are often desired for techniques or for targeted gene correction, Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000). For double stranded RNA inhibition which may utilize either natural ds RNA or ds RNA modified in its, sugar, phosphate or base, *see* Hannon, *Nature* 418(11): 244-251 (2002); Fire *et al.* in WO 99/32619; Tuschl *et al.* in US2002/0086356; Kruetzer *et al.* in WO 00/44895, the disclosures of which are incorporated herein by reference in their entirety;. For circular antisense, see Kool in U.S. Patent No. 5,426,180, the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative U.S. Patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U.S. Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred nucleic acid molecules, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference in its entirety. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA). PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex

(in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999).

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), and U.S. Patent Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acid molecules of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); and Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994). Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.*

130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997); Rowley *et al.*, *Mol Med* 5(10): 693-700 (1999); Kool, *Annu Rev Biophys Biomol Struct.* 25: 1-28 (1996).

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect and characterize gross alterations in the gene of a BSNA, such as deletions, insertions, translocations, and duplications of the BSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999). The isolated nucleic acid molecules of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include a nucleic acid molecule of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level. Alternatively, detection techniques such as molecular beacons may be used, see Kostrikis *et al. Science* 279:1228-1229 (1998).

The isolated nucleic acid molecules of the present invention can be also be used as probes to detect, characterize, and quantify BSNA in, and isolate BSNA from, transcript-derived nucleic acid samples. In one embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^{+}- selected RNA samples. In another embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000). In another preferred embodiment, the

isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to BSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000).

In another embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify and/or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In this embodiment, it is preferred that the probe or primer be derived from a nucleic acid molecule encoding a BSP. More preferably, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 95-156. Also preferred are probes or primers derived from a BSNA. More preferred are probes or primers derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-94.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*,

in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); and McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995). Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; and Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995).

PCR and hybridization methods may be used to identify and/or isolate nucleic acid molecules of the present invention including allelic variants, homologous nucleic acid molecules and fragments. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules of the present invention that encode homologous proteins, analogs, fusion protein or muteins of the invention. Nucleic acid primers as described herein can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

These nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); international patent publications WO 97/19193 and WO 00/15779, and U.S. Patent Nos. 5,854,033 and 5,714,320, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that comprise one or more of the nucleic acid molecules of the present invention.

In yet another embodiment, the invention is directed to single exon probes based on the BSNAs disclosed herein.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention provides vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acid molecules of the present invention in host cells (cloning vectors), for shuttling the nucleic acid molecules of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acid molecules of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acid molecules of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acid molecules of the present invention, alone or as fusion proteins with heterologous polypeptides (expression vectors). Vectors are by now well known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*. Furthermore, a variety of vectors are available commercially. Use of existing vectors and modifications thereof are well within the skill in the art. Thus, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences that control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the

nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells may be chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda, e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3'

of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

The host cells may also be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include, include but are not limited to, resistance to neomycin (G418), blasticidin, hygromycin and zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention

may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the nucleic acid molecules of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, and the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast $\underline{\alpha}$-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include, but are not limited to, those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-

promoter from SV40 and the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the BSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows a high level expression of an RNA that encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response

element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one embodiment of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Such tags include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the polypeptides of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope, detectable by anti-HA antibody.

For secretion of expressed polypeptides, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or

identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α-agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999). A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced

GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); U.S. Patent Nos. 6,090,919 and 5,804,387, the disclosures of which are incorporated herein by reference in their entireties) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999); Yang, *et al., J Biol Chem*, 273: 8212-6 (1998); Bevis *et al.*, *Nature Biotechnology*, 20:83-7 (2002). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half-life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412, the disclosures of which are incorporated herein by reference in their entireties.

For long-term, high-yield recombinant production of the polypeptides of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of

recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA) allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid molecules of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed polypeptide in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide BSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid molecules of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion

characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid molecules of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid molecules according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods that are well known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella*

typhimurium; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda* — *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA) — *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from breast are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human breast cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See*, *e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*.

Methods for introducing the vectors and nucleic acid molecules of the present invention into the host cells are well known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli,* in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^+ or K^+, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent

strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent to take up exogenous DNA by electroporation by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided by BioRad (Richmond, CA, USA).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as a snail-gut extract, usually denoted Glusulase or Zymolyase, or an enzyme from *Arthrobacter luteus* to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate to permeabilize the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian

Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found in, for example, ; Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000). Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al., Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well within the skill in the art and thus need not be detailed here. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001).

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides, including Fragments Muteins, Homologous Proteins, Allelic Variants, Analogs and Derivatives

Another aspect of the invention relates to polypeptides encoded by the nucleic acid molecules described herein. In a preferred embodiment, the polypeptide is a breast

specific polypeptide (BSP). In an even more preferred embodiment, the polypeptide comprises an amino acid sequence of SEQ ID NO:95-156 or is derived from a polypeptide having the amino acid sequence of SEQ ID NO: 95-156. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well known to those having ordinary skill in the art.

Polypeptides of the present invention may also comprise a part or fragment of a BSP. In a preferred embodiment, the fragment is derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 95-156. Polypeptides of the present invention comprising a part or fragment of an entire BSP may or may not be BSPs. For example, a full-length polypeptide may be breast-specific, while a fragment thereof may be found in other tissues as well as in breast. A polypeptide that is not a BSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-BSP antibodies. In a preferred embodiment, the part or fragment is a BSP. Methods of determining whether a polypeptide of the present invention is a BSP are described *infra*.

Polypeptides of the present invention comprising fragments of at least 6 contiguous amino acids are also useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of a polypeptide of the present invention have utility in such a study.

Polypeptides of the present invention comprising fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize polypeptides of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983). As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic and are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the polypeptides of the present invention have utility as immunogens.

Polypeptides comprising fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire polypeptide, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the polypeptide of interest. See U.S. Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The polypeptide of the present invention thus preferably is at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the polypeptide of the present invention is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger polypeptides having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments by truncating the nucleic acid molecule, *e.g.*, a BSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally occurring polypeptide. Methods of producing polypeptide fragments are well known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment, preferably a fragment of a BSP, may be produced by chemical or enzymatic cleavage of a BSP polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule of the present invention encoding a fragment, preferably of a BSP, in a host cell.

Polypeptides of the present invention are also inclusive of mutants, fusion proteins, homologous proteins and allelic variants.

A mutant protein, or mutein, may have the same or different properties compared to a naturally occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native polypeptide. Small deletions and insertions can often be found that do not alter the function of a protein. Muteins may or may not be breast-specific. Preferably, the mutein is breast-specific. More preferably the mutein is a polypeptide that comprises at

least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 95-156. Accordingly, in a preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 95-156. In a yet more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 95-156.

A mutein may be produced by isolation from a naturally occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein is produced from a host cell comprising a mutated nucleic acid molecule compared to the naturally occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid molecule of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is breast-specific, as described below. Multiple random mutations can be introduced into the gene by methods well known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), as well as U.S. Patent No. 5,223,408, which is herein incorporated by reference in its entirety.

The invention also contemplates polypeptides that are homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to a BSP. In an even more preferred embodiment, the polypeptide is homologous to a

BSP selected from the group having an amino acid sequence of SEQ ID NO: 95-156. By homologous polypeptide it is means one that exhibits significant sequence identity to a BSP, preferably a BSP having an amino acid sequence of SEQ ID NO: 95-156. By significant sequence identity it is meant that the homologous polypeptide exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 95-156. More preferred are homologous polypeptides exhibiting at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 95-156. Most preferably, the homologous polypeptide exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 95-156. In a preferred embodiment, the amino acid substitutions of the homologous polypeptide are conservative amino acid substitutions as discussed above.

Homologous polypeptides of the present invention also comprise polypeptide encoded by a nucleic acid molecule that selectively hybridizes to a BSNA or an antisense sequence thereof. In this embodiment, it is preferred that the homologous polypeptide be encoded by a nucleic acid molecule that hybridizes to a BSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. More preferred is a homologous polypeptide encoded by a nucleic acid sequence which hybridizes to a BSNA selected from the group consisting of SEQ ID NO: 1-94 or a homologous polypeptide encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a BSP, preferably an BSP of SEQ ID NO:95-156 under low stringency, moderate stringency or high stringency conditions, as defined herein.

Homologous polypeptides of the present invention may be naturally occurring and derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, or baboon, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 95-156. The homologous polypeptide may also be a naturally occurring polypeptide from a human, when the BSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*,

dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. The homologous polypeptide may also be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. Alternatively, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a BSP. In a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a BSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated polypeptide not only identical in sequence to those described with particularity herein, but also to provide isolated polypeptide ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, polypeptides of the present invention are also inclusive of those encoded by an allelic variant of a nucleic acid molecule encoding a BSP. In this embodiment, it is preferred that the polypeptide be encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 95-156. More preferred is that the polypeptide be encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-94.

Polypeptides of the present invention are also inclusive of derivative polypeptides encoded by a nucleic acid molecule according to the instant invention. In this embodiment, it is preferred that the polypeptide be a BSP. Also preferred are derivative polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NO: 95-156 and which has been acetylated, carboxylated, phosphorylated,

glycosylated, ubiquitinated or other PTMs. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed November 11, 2002), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications include, but are not limited to: (Z)-dehydrobutyrine; 1-chondroitin sulfate-L-aspartic acid ester; 1'-glycosyl-L-tryptophan; 1'-phospho-L-histidine; 1-thioglycine; 2'-(S-L-cysteinyl)-L-histidine; 2'-[3-carboxamido (trimethylammonio)propyl]-L-histidine; 2'-alpha-mannosyl-L-tryptophan; 2-methyl-L-glutamine; 2-oxobutanoic acid; 2-pyrrolidone carboxylic acid; 3'-(1'-L-histidyl)-L-tyrosine; 3'-(8alpha-FAD)-L-histidine; 3'-(S-L-cysteinyl)-L-tyrosine; 3', 3",5'-triiodo-L-

thyronine; 3'-4'-phospho-L-tyrosine; 3-hydroxy-L-proline; 3'-methyl-L-histidine; 3-methyl-L-lanthionine; 3'-phospho-L-histidine; 4'-(L-tryptophan)-L-tryptophyl quinone; 42 N-cysteinyl-glycosylphosphatidylinositolethanolamine; 43 -(T-L-histidyl)-L-tyrosine; 4-hydroxy-L-arginine; 4-hydroxy-L-lysine; 4-hydroxy-L-proline; 5'-(N6-L-lysine)-L-topaquinone; 5-hydroxy-L-lysine; 5-methyl-L-arginine; alpha-l-microglobulin-Ig alpha complex chromophore; bis-L-cysteinyl bis-L-histidino diiron disulfide; bis-L--cysteinyl-L-N3'-histidino-L-serinyl tetrairon' tetrasulfide; chondroitin sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-alanine; D-allo-isoleucine; D-asparagine; dehydroalanine; dehydrotyrosine; dermatan 4-sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-glucuronyl-N-glycine; dipyrrolylmethanemethyl-L-cysteine; D-leucine; D-methionine; D-phenylalanine; D-serine; D-tryptophan; glycine amide; glycine oxazolecarboxylic acid; glycine thiazolecarboxylic acid; heme P450-bis-L-cysteine-L-tyrosine; heme-bis-L-cysteine; hemediol-L-aspartyl ester-L-glutamyl ester; hemediol-L-aspartyl ester-L-glutamyl ester-L-methionine sulfonium; heme-L-cysteine; heme-L-histidine; heparan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; heme P450-bis-L-cysteine-L-lysine; hexakis-L-cysteinyl hexairon hexasulfide; keratan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-threonine; L oxoalanine- lactic acid; L phenyllactic acid; l'-(8alpha-FAD)-L-histidine; L-2'.4',5'-topaquinone; L-3',4'-dihydroxyphenylalanine; L-3'.4'.5'-trihydroxyphenylalanine; L-4'-bromophenylalanine; L-6'-bromotryptophan; L-alanine amide; L-alanyl imidazolinone glycine; L-allysine; L-arginine amide; L-asparagine amide; L-aspartic 4-phosphoric anhydride; L-aspartic acid 1-amide; L-beta-methylthioaspartic acid; L-bromohistidine; L-citrulline; L-cysteine amide; L-cysteine glutathione disulfide; L-cysteine methyl disulfide; L-cysteine methyl ester; L-cysteine oxazolecarboxylic acid; L-cysteine oxazolinecarboxylic acid; L-cysteine persulfide; L-cysteine sulfenic acid; L-cysteine sulfinic acid; L-cysteine thiazolecarboxylic acid; L-cysteinyl homocitryl molybdenum-heptairon-nonasulfide; L-cysteinyl imidazolinone glycine; L-cysteinyl molybdopterin; L-cysteinyl molybdopterin guanine dinucleotide; L-cystine; L-erythro-beta-hydroxyasparagine; L-erythro-beta-hydroxyaspartic acid; L-gamma-carboxyglutamic acid; L-glutamic acid 1-amide; L-glutamic acid 5-methyl ester; L-glutamine amide; L-glutamyl 5-glycerylphosphorylethanolarnine; L-histidine amide; L-isoglutamyl-polyglutamic acid; L-isoglutamyl-polyglycine; L-isoleucine amide; L-lanthionine; L-leucine amide; L-lysine amide; L-lysine thiazolecarboxylic acid; L-lysinoalanine; L-methionine amide; L-

methionine sulfone; L-phenylalanine thiazolecarboxylic acid; L-phenylalanine amide; L-proline amide; L-selenocysteine; L-selenocysteinyl molybdopterin guanine dinucleotide; L-serine amide; L-serine thiazolecarboxylic acid; L-seryl imidazolinone glycine; L-T-bromophenylalanine; L-T-bromophenylalanine; L-threonine amide; L-thyroxine; L-tryptophan amide; L-tryptophyl quinone; L-tyrosine amide; L-valine amide; meso-lanthionine; N-(L-glutamyl)-L-tyrosine; N-(L-isoaspartyl)-glycine; N-(L-isoaspartyl)-L-cysteine; N,N,N-trimethyl-L-alanine; N,N-dimethyl-L-proline; N2-acetyl-L-lysine; N2-succinyl-L-tryptophan; N4-(ADP-ribosyl)-L-asparagine; N4-glycosyl-L-asparagine; N4-hydroxymethyl-L-asparagine; N4-methyl-L-asparagine; N5-methyl-L-glutamine; N6- 1 -carboxyethyl-L-lysine; N6-(4-amino hydroxybutyl)-L-lysine; N6-(L-isoglutamyl)-L-lysine; N6-(phospho-5'-adenosine)-L-lysine; N6-(phospho-5'-guanosine)-L-tysine; N6,N6,N6-trimethyl-L-lysine; N6,N6-dimethyl-L-lysine; N6-acetyl-L-lysine; N6-biotinyl-L-lysine; N6-carboxy-L-lysine; N6-formyl-L-lysine; N6-glycyl-L-lysine; N6-lipoyl-L-lysine; N6-methyl-L-lysine; N6-methyl-N6-poly(N-methyl-propylamine)-L-lysine; N6-mureinyl-L-lysine; N6-myristoyl-L-lysine; N6-palmitoyl-L-lysine; N6-pyridoxal phosphate-L-lysine; N6-pyruvic acid 2-iminyl-L-lysine; N6-retinal-L-lysine; N-acetylglycine; N-acetyl-L-glutamine; N-acetyl-L-alanine; N-acetyl-L-aspartic acid; N-acetyl-L-cysteine; N-acetyl-L-glutamic acid; N-acetyl-L-isoleucine; N-acetyl-L-methionine; N-acetyl-L-proline; N-acetyl-L-serine; N-acetyl-L-threonine; N-acetyl-L-tyrosine; N-acetyl-L-valine; N-alanyl-glycosylphosphatidylinositolethanolamine; N-asparaginyl-glycosylphosphatidylinositolethanolarnine; N-aspartyl-glycosylphosphatidylinositolethanolamine; N-formylglycine; N-formyl-L-methionine; N-glycyl-glycosylphosphatidylinositolethanolamine; N-L-glutamyl-poly-L-glutamic acid; N-methylglycine; N-methyl-L-alanine; N-methyl-L-methionine; N-methyl-L-phenylalanine; N-myristoyl-glycine; N-palmitoyl-L-cysteine; N-pyruvic acid 2-iminyl-L-cysteine; N-pyruvic acid 2-iminyl-L-valine; N-seryl-glycosylphosphatidylinositolethanolamine; N-seryl-glycosyBSPhingolipidinositolethanolamine; O-(ADP-ribosyl)-L-serine; O-(phospho-5'-adenosine)-L-threonine; O-(phospho-5'-DNA)-L-serine; O-(phospho-5'-DNA)-L-threonine; O-(phospho-5'rRNA)-L-serine; O-(phosphoribosyl dephospho-coenzyme A)-L-serine; O-(sn-l-glycerophosphoryl)-L-serine; O4'-(8alpha-FAD)-L-tyrosine; O4'-(phospho-5'-adenosine)-L-tyrosine; O4'-(phospho-5'-DNA)-L-tyrosine; O4'-(phospho-5'-RNA)-L-tyrosine; O4'-(phospho-5'-uridine)-L-tyrosine; O4-glycosyl-L-hydroxyproline; O4'-glycosyl-L-tyrosine; O4'-sulfo-L-tyrosine; O5-glycosyl-L-hydroxylysine; O-glycosyl-L-

serine; O-glycosyl-L-threonine; omega-N-(ADP-ribosyl)-L-arginine; omega-N-omega-N'-dimethyl-L-arginine; omega-N-methyl-L-arginine; omega-N-omega-N-dimethyl-L-arginine; omega-N-phospho-L-arginine; O'octanoyl-L-serine; O-palmitoyl-L-serine; O-palmitoyl-L-threonine; O-phospho-L-serine; O-phospho-L-threonine; O-phosphopantetheine-L-serine; phycoerythrobilin-bis-L-cysteine; phycourobilin-bis-L-cysteine; pyrroloquinoline quinone; pyruvic acid; S hydroxycinnamyl-L-cysteine; S-(2-aminovinyl) methyl-D-cysteine; S-(2-aminovinyl)-D-cysteine; S-(6-FW-L-cysteine; S-(8alpha-FAD)-L-cysteine; S-(ADP-ribosyl)-L-cysteine; S-(L-isoglutamyl)-L-cysteine; S-12-hydroxyfarnesyl-L-cysteine; S-acetyl-L-cysteine; S-diacylglycerol-L-cysteine; S-diphytanylglycerot diether-L-cysteine; S-farnesyl-L-cysteine; S-geranylgeranyl-L-cysteine; S-glycosyl-L-cysteine; S-glycyl-L-cysteine; S-methyl-L-cysteine; S-nitrosyl-L-cysteine; S-palmitoyl-L-cysteine; S-phospho-L-cysteine; S-phycobiliviolin-L-cysteine; S-phycocyanobilin-L-cysteine; S-phycoerythrobilin-L-cysteine; S-phytochromobilin-L-cysteine; S-selenyl-L-cysteine; S-sulfo-L-cysteine; tetrakis-L-cysteinyl diiron disulfide; tetrakis-L-cysteinyl iron; tetrakis-L-cysteinyl tetrairon tetrasulfide; trans-2,3-cis 4-dihydroxy-L-proline; tris-L-cysteinyl triiron tetrasulfide; tris-L-cysteinyl triiron trisulfide; tris-L-cysteinyl-L-aspartato tetrairon tetrasulfide; tris-L-cystêinyl-L-cysteine persulfido-bis-L-glutamato-L-histidino tetrairon disulfide trioxide; tris-L-cysteinyl-L-N3'-histidino tetrairon tetrasulfide; tris-L-cysteinyl-L-Nl'-histidino tetrairon tetrasulfide; and tris-L-cysteinyl-L-serinyl tetrairon tetrasulfide.

Additional examples of PTMs may be found in web sites such as the Delta Mass database based on Krishna, R. G. and F. Wold (1998). Posttranslational Modifications. Proteins - Analysis and Design. R. H. Angeletti. San Diego, Academic Press. 1: 121-206. ; Methods in Enzymology, 193, J.A. McClosky (ed) (1990), pages 647-660; Methods in Protein Sequence Analysis edited by Kazutomo Imahori and Fumio Sakiyama, Plenum Press, (1993) "Post-translational modifications of proteins" R.G. Krishna and F. Wold pages 167-172; "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) “O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins” Gupta et al. Nucleic Acids Research, 27: 370-372 (1999); and "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al.Nucleic Acids Res 27(1):237-239 (1999) see also, WO 02/21139A2, the disclosure of which is incorporated herein by reference in its entirety.

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signalling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is

cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-

translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule may also be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslational events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430,

Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

Polypeptides of the present invention, including full length polypeptides, fragments and fusion proteins, can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to polypeptides of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-BSP antibodies.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999). PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

Polypeptides of the present invention are also inclusive of analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, this polypeptide is a BSP. In a more preferred embodiment, this polypeptide is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 95-156. Also preferred is an analog polypeptide comprising one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally occurring polypeptide. In one embodiment, the analog is structurally similar to a BSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --CH(OH)CH_2-- and –CH_2SO--. In another embodiment, the analog comprises substitution of one or more amino acids of a BSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see,*

e.g., Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993).

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl--(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-

aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

Another aspect of the present invention relates to the fusion of a polypeptide of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide of the present invention is a BSP. In a more preferred embodiment, the polypeptide of the present invention that is fused to a heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 95-156, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the fusion protein is encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NO: 1-94, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94.

The fusion proteins of the present invention will include at least one fragment of a polypeptide of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the polypeptide of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of a polypeptide of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably at least 15, 20, or 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included

render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins into the periplasmic space or extracellular milieu for prokaryotic hosts or into the culture medium for eukaryotic cells through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful fusion proteins of the present invention include those that permit use of the polypeptide of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); Colas *et al.*, *Nature* 380, 548-550 (1996); Norman, T. *et al.*, *Science* 285, 591-595 (1999); Fabbrizio *et al.*, *Oncogene* 18, 4357-4363 (1999); Xu *et al.*, *Proc Natl Acad Sci U S A.* 94, 12473-12478 (1997); Yang, *et al.*, *Nuc. Acids Res.* 23, 1152-1156 (1995); Kolonin *et al.*, *Proc Natl Acad Sci U S A* 95, 14266-14271 (1998); Cohen *et al.*, *Proc Natl Acad Sci U S A* 95, 14272-14277 (1998); Uetz, *et al. Nature* 403, 623-627(2000); Ito, *et al.*, *Proc Natl Acad Sci U S A* 98, 4569-4574 (2001). Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded polypeptide on the surface of a phage or cell, fusions to intrinsically fluorescent proteins,

such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above.

The polypeptides of the present invention can also usefully be fused to protein toxins, such as Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the BSP.

As further described below, the polypeptides of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize polypeptides of the present invention including BSPs and their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly BSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of BSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of BSPs.

One may determine whether polypeptides of the present invention including BSPs, muteins, homologous proteins or allelic variants or fusion proteins of the present invention

are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the polypeptide at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, (Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides or fusion proteins of the present invention is well known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated polypeptides or fusion proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous and non-proteinaceous material and are well known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated polypeptide or fusion protein of the present invention are used as therapeutic agents, such as in vaccines and replacement therapy, the isolated polypeptides of the present invention are also useful at lower purity. For example, partially purified polypeptides of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In a preferred embodiment, the purified and substantially purified polypeptides of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides or fusion proteins of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent. For example, the peptides of the invention may be

stabilized by covalent linkage to albumin. See, U.S. Patent No. 5,876,969, the contents of which are hereby incorporated in its entirety.

For example, the polypeptides or fusion proteins of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the polypeptides or fusion proteins of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized polypeptide or fusion protein of the present invention.

As another example, the polypeptides or fusion proteins of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides and fusion proteins of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biologic interaction there between. The polypeptides or fusion proteins of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biological interaction there between.

Alternative Transcripts

In antother aspect, the present invention provides splice variants of genes and proteins encoded thereby. The identification of a novel splice varaint which encodes an amino acid sequence with a novel region can be targeted for the generation of reagents for use in detection and/or treatment of cancer. The novel amino acid sequence may lead to a unique protein structure, protein subcellular localization, biochemical processing or

function of the splice varaint. This information can be used to directly or indirectly facilitate the generation of additional or novel therapeutics or diagnostics. The nucleotide sequence in this novel splice variant can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

Specifically, the newly identified sequences may enable the production of new antibodies or compounds directed against the novel region for use as a therapeutic or diagnostic. Alternatively, the newly identified sequences may alter the biochemical or biological properties of the encoded protein in such a way as to enable the generation of improved or different therapeutics targeting this protein.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention. In a preferred embodiment, the antibodies are specific for a polypeptide that is a BSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 95-156, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a BSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or vis versa. In addition, alternative splice forms of a BSP may be indicative of cancer. Differential degradation of the C or N-terminus of a BSP may also be a marker or target for anticancer therapy. For example, an BSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-BSP polypeptides by at least two-fold,

more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the polypeptide of the present invention in samples derived from human breast.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to $1 \mathrm{X} 10^{-13}$ M proving especially useful.

The antibodies of the present invention can be naturally occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the polypeptides of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the polypeptide of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will

often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster), lagomorphs (typically rabbits), and also larger mammals, such as sheep, goats, cows, and horses; or egg laying birds or reptiles such as chickens or alligators. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the polypeptide of the present invention. One form of avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of a polypeptide of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide of the present invention to other moieties. For example, polypeptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), <u>Using Antibodies: A Laboratory Manual</u>, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), <u>Current Protocols in Immunology</u>, John Wiley & Sons, Inc. (2001); Zola, <u>Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench)</u>, Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996). Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the polypeptides of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the polypeptides of the present invention. Antibodies from avian species may have particular advantage in detection of the polypeptides of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998). Following immunization, the antibodies of the present invention can be obtained using any art-accepted technique. Such techniques are well known in the art and are described in detail in references such as Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), <u>Basic Methods in Antibody Production and Characterization</u>, CRC Press (2000); Harlow, *supra*; Davis (ed.), <u>Monoclonal Antibody Protocols</u>, Vol. 45, Humana Press (1995); Delves (ed.), <u>Antibody Production: Essential Techniques</u>, John Wiley & Son Ltd (1997); and Kenney, <u>Antibody Solution: An Antibody Methods Manual</u>, Chapman & Hall (1997).

Briefly, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the polypeptides of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the polypeptides of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant antibody production of whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established.

See, e.g., Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas (2001), *supra*; Kay, *supra*; and Abelson, *supra.*

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell. Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention. For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae. See, e.g.,* Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); and Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992).

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); and Limonta et al., *Immunotechnology* 1: 107-13 (1995).

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells. Verma *et al., J. Immunol. Methods* 216(1-2):165-81 (1998) review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies. Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al., J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al., Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al., J. Immunol. Methods* 231(1-2): 147-57 (1999).

The invention further provides antibody fragments that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. Among such useful fragments are Fab, Fab', Fv, $F(ab')_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

The present invention also relates to antibody derivatives that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus are more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful method is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985); and U.S. Patent No. 5,807,715 the disclosure of which is incorporated herein by reference in its entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); and U.S. Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties. Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. Accordingly, the present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., Proc. Natl. Acad. Sci. (USA) 90: 7889-7893 (1993); Duan et al., Proc. Natl. Acad. Sci. (USA) 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited

by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product. Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See*, *e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995). Kits for such enhanced chemiluminescent detection (ECL) are available commercially. The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores. There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I. As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the polypeptides of the present invention. Commonly, the

antibody in such immunotoxins is conjugated to Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998).

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, attached to a substrate. Substrates can be porous or nonporous, planar or nonplanar. For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography. For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the polypeptides of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the BSPs of the present invention or to polypeptides encoded by the BSNAs of the invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a BSP. In a preferred embodiment, the BSP comprises an amino acid sequence selected from SEQ ID NO: 95-156, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a BSNA of the invention, preferably a BSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-94, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human BSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent No. 4,873,191, herein incorporated by reference in its entirety); retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.*, Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.*, Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.*, Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.*, Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.*, Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232- 6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.,* Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See*, *e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a

patient in vivo. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.,* U.S. Patent Nos. 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 95-156 and SEQ ID NO: 1-94 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set

representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence. In addition, the invention includes a method of using patterns of expression associated with either the nucleic acids or proteins in a computer-based method to diagnose disease.

Diagnostic Methods for breast Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a BSNA or a BSP in a human patient that has or may have breast

cancer, or who is at risk of developing breast cancer, with the expression of a BSNA or a BSP in a normal human control. For purposes of the present invention, "expression of a BSNA" or "BSNA expression" means the quantity of BSNA mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a BSP" or "BSP expression" means the amount of BSP that can be measured by any method known in the art or the level of translation of a BSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing breast cancer in a patient, by analyzing for changes in levels of BSNA or BSP in cells, tissues, organs or bodily fluids compared with levels of BSNA or BSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a BSNA or BSP in the patient versus the normal human control is associated with the presence of breast cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in the structure of the mRNA of a BSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in a BSP compared to a BSP from a normal patient. These changes include, *e.g.*, alterations, including post translational modifications such as glycosylation and/or phosphorylation of the BSP or changes in the subcellular BSP localization.

For purposes of the present invention, diagnosing means that BSNA or BSP levels are used to determine the presence or absence of disease in a patient. As will be understood by those of skill in the art, measurement of other diagnostic parameters may be required for definitive diagnosis or determination of the appropriate treatment for the disease. The determination may be made by a clinician, a doctor, a testing laboratory, or a patient using an over the counter test. The patient may have symptoms of disease or may be asymptomatic. In addition, the BSNA or BSP levels of the present invention may be used as screening marker to determine whether further tests or biopsies are warranted. In addition, the BSNA or BSP levels may be used to determine the vulnerability or susceptibility to disease.

In a preferred embodiment, the expression of a BSNA is measured by determining the amount of a mRNA that encodes an amino acid sequence selected from SEQ ID NO: 95-156, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the BSNA expression that is measured is the level of expression of a BSNA mRNA selected from SEQ ID NO: 1-94, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acid molecules. BSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. BSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a BSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, BSNA expression may be compared to a known control, such as normal breast nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a BSP is measured by determining the level of a BSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 95-156, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of a BSNA or BSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of breast cancer. The expression level of a BSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the BSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the BSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine,

phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a BSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-BSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the BSP will bind to the anti-BSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-BSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the BSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an BSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure BSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-BSP antibody is attached to a solid support and an allocated amount of a labeled BSP and a sample of interest are incubated with the solid support. The amount of labeled BSP attached to the solid support can be correlated to the quantity of a BSP in the sample.

Of the proteomic approaches, 2D PAGE is a well known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a BSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more BSNAs of interest. In this approach, all or a portion of one or more BSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. As used herein "blood" includes whole blood, plasma, serum, circulating epithelial cells, constituents, or any derivative of blood.

In addition to detection in bodily fluids, the proteins and nucleic acids of the invention are suitable to detection by cell capture technology. Whole cells may be captured by a variety methods for example magnetic separation, U.S. Patent. Nos. 5,200,084; 5,186,827; 5,108,933; 4,925,788, the disclosures of which are incorporated herein by reference in their entireties. Epithelial cells may be captured using such products as Dynabeads® or CELLection™ (Dynal Biotech, Oslo, Norway). Alternatively, fractions of blood may be captured, e.g., the buffy coat fraction (50mm cells isolated from

5ml of blood) containing epithelial cells. In addition, cancer cells may be captured using the techniques described in WO 00/47998, the disclosure of which is incorporated herein by reference in its entirety. Once the cells are captured or concentrated, the proteins or nucleic acids are detected by the means described in the subject application. Alternatively, nucleic acids may be captured directly from blood samples, see U.S. Patent Nos. 6,156,504, 5,501,963; or WO 01/42504 , the disclosures of which are incorporated herein by reference in their entireties.

In a preferred embodiment, the specimen tested for expression of BSNA or BSP includes without limitation breast tissue, breast cells grown in cell culture, blood, serum, lymph node tissue, and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary breast cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, lungs, colon, and adrenal glands. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a BSNA or BSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other BSNA or BSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular BSNA or BSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more BSNA and/or BSP in a sample from a patient suspected of having breast cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP and then ascertaining whether the patient has breast

cancer from the expression level of the BSNA or BSP. In general, if high expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether breast cancer has metastasized in a patient. One may identify whether the breast cancer has metastasized by measuring the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a variety of tissues. The presence of a BSNA or BSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a BSNA or BSP is associated with breast cancer. Similarly, the presence of a BSNA or BSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a BSNA or BSP is associated with breast cancer. Further, the presence of a structurally altered BSNA or BSP that is associated with breast cancer is also indicative of metastasis.

In general, if high expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten

times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

Staging

The invention also provides a method of staging breast cancer in a human patient. The method comprises identifying a human patient having breast cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more BSNAs or BSPs. First, one or more tumors from a variety of patients are staged according to procedures well known in the art, and the expression levels of one or more BSNAs or BSPs is determined for each stage to obtain a standard expression level for each BSNA and BSP. Then, the BSNA or BSP expression levels of the BSNA or BSP are determined in a biological sample from a patient whose stage of cancer is not known. The BSNA or BSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the BSNAs and BSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a BSNA or BSP to determine the stage of a breast cancer.

Monitoring

Further provided is a method of monitoring breast cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the breast cancer. The monitoring may determine if there has been a reoccurrence and, if so, determine its nature. The method comprises identifying a human patient that one wants to monitor for breast cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more BSNAs or BSPs, and comparing the BSNA or BSP levels over time to those BSNA or BSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a BSNA or BSP that are associated with breast cancer.

If increased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an

increase in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting a decrease in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of BSNAs or BSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of breast cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a BSNA and/or BSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more BSNAs and/or BSPs are detected. The presence of higher (or lower) BSNA or BSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly breast cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more BSNAs and/or BSPs of the invention can also be monitored by analyzing levels of expression of the BSNAs and/or BSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a BSG, thereby determining if a human with the genetic lesion is susceptible to developing breast cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing breast cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the BSGs of this invention, a chromosomal rearrangement

of a BSG, an aberrant modification of a BSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a BSG. Methods to detect such lesions in the BSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous breast Diseases

The present invention also provides methods for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having or known to have a noncancerous breast disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP, comparing the expression level or structural alteration of the BSNA or BSP to a normal breast control, and then ascertaining whether the patient has a noncancerous breast disease. In general, if high expression relative to a control of a BSNA or BSP is indicative of a particular noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of a noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a BSNA and/or BSP is associated with a particular noncancerous breast disease by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining which BSNAs and/or BSPs are expressed in the tissue at either a higher or a lower level than in normal breast tissue. In another embodiment, one may determine whether a BSNA or BSP exhibits structural alterations in a particular noncancerous breast disease state by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining the structural alterations in one or more BSNAs and/or BSPs relative to normal breast tissue.

Methods for Identifying breast Tissue

In another aspect, the invention provides methods for identifying breast tissue. These methods are particularly useful in, *e.g.*, forensic science, breast cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is breast tissue or has breast tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising breast tissue or having breast tissue-like characteristics, determining whether the sample expresses one or more BSNAs and/or BSPs, and, if the sample expresses one or more BSNAs and/or BSPs, concluding that the sample comprises breast tissue. In a preferred embodiment, the BSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 95-156, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from SEQ ID NO: 1-94, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a BSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a BSP is expressed. Determining whether a sample expresses a BSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the BSP has an amino acid sequence selected from SEQ ID NO: 95-156, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two BSNAs and/or BSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five BSNAs and/or BSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is breast tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into breast tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new breast tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation

include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying breast Tissue

In another aspect, the invention provides methods for producing engineered breast tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a BSNA or a BSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of breast tissue cells. In a preferred embodiment, the cells are pleuripotent. As is well known in the art, normal breast tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered breast tissue or cells comprises one of these cell types. In another embodiment, the engineered breast tissue or cells comprises more than one breast cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the breast cell tissue. Methods for manipulating culture conditions are well known in the art.

Nucleic acid molecules encoding one or more BSPs are introduced into cells, preferably pleuripotent cells. In a preferred embodiment, the nucleic acid molecules encode BSPs having amino acid sequences selected from SEQ ID NO: 95-156, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1-94, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a BSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well known in the art and are described in detail, *supra*.

Artificial breast tissue may be used to treat patients who have lost some or all of their breast function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, fusion proteins, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, or inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a BSNA or part thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-94, a nucleic acid that hybridizes thereto, an allelic

variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a BSP or fragment thereof. In a more preferred embodiment, the pharmaceutical composition comprises a BSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 95-156, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-BSP antibody, preferably an antibody that specifically binds to a BSP having an amino acid that is selected from the group consisting of SEQ ID NO: 95-156, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art that is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and

tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, cornstarch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically. For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example BSP polypeptide, fusion protein, or fragments thereof, antibodies specific for BSP, agonists, antagonists or inhibitors of BSP, which ameliorates the signs or symptoms of the disease or prevent progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of breast function. As used herein, “treating” includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for the purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further described in U.S. Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891;

5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid molecule of the present invention is administered. The nucleic acid molecule can be delivered in a vector that drives expression of a BSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a BSP are administered, for example, to complement a deficiency in the native BSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a BSP having the amino acid sequence of SEQ ID NO: 95-156, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a BSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in BSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a BSP having the amino acid sequence of SEQ ID NO: 95-156, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a BSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a BSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a BSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to BSG transcripts, are also useful in therapy. *See, e.g.*,

Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995).

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the BSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); and McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000). Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a BSP, preferably a BSP comprising an amino acid sequence of SEQ ID NO: 95-156, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-94, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a BSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant BSP defect.

Protein compositions are administered, for example, to complement a deficiency in native BSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to BSP. The immune response can be used to modulate activity of BSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate BSP.

In a preferred embodiment, the polypeptide administered is a BSP comprising an amino acid sequence of SEQ ID NO: 95-156, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded

by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-94, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well known, antibody compositions are administered, for example, to antagonize activity of BSP, or to target therapeutic agents to sites of BSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a BSP comprising an amino acid sequence of SEQ ID NO: 95-156, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-94, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a BSP or have a modulatory effect on the expression or activity of a BSP. Modulators which decrease the expression or activity of BSP (antagonists) are believed to be useful in treating breast cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a BSP can also be designed, synthesized and tested for use in the imaging and treatment of breast cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the BSPs identified herein. Molecules identified in the library as being capable of binding to a BSP are key candidates for further evaluation for use in the treatment of breast cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a BSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of BSP is administered. Antagonists of BSP can be produced using methods generally known in the art. In particular, purified BSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a BSP.

In other embodiments a pharmaceutical composition comprising an agonist of a BSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP comprising an amino acid sequence of SEQ ID NO: 95-156, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-94, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting breast Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the breast or to specific cells in the breast. In a preferred embodiment, an anti-BSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if breast tissue needs to be selectively destroyed. This would be useful for targeting and killing breast cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting breast cell function.

In another embodiment, an anti-BSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring breast function, identifying breast cancer tumors, and identifying noncancerous breast diseases.

EXAMPLES

Example 1: Gene Expression analysis

Identification of BSGs was carried out by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc, Palo Alto, CA, using the data mining software package CLASP™.

The CLASP target gene identification process is focused on, but not limited to, the following 4 CLASP profiles: tissue specific expression, cancer specific expression, differentially expressed in cancer, maximum tissue differential expression.

(1) For these profiles:cDNA libraries were divided into 48 unique tissue organs. The genes were grouped into gene bins, each bin is a sequence based cluster grouped together with a common contig. The expression levels for each gene bin were calculated in each organ. Differential expression significance was calculated with rigorous statistical significant test considering the influence of sequence random fluctuations and sampling size of cDNA libraries from concept published by Audic S and Claverie JM (Genome Res 1997 7(10): 986-995: The significance of digital gene expression profiles).

(2) Highly expressed organ specific genes were selected based on the percentage abundance level in the targeted organ versus all the other organs (organ-specificity).

(3) The expression levels of each highly expressed organ-specific gene in the tumor tissue libraries were compared with normal tissue libraries and tissue libraries associated with tumor or disease (cancer-specificity) and analyzed for statistical significance.

(4) Target genes exhibiting each CLASP profile criteria were selected

CLASP tissue specific expression profile: In order to meet the organ-specificity criteria, the expression level of the component clones which the gene is composed of must exhibit 3 or more occurrences regardless the total number of genes isolated for the target organ. The percentage abundance level in each organ was calculated to identify the organ with the highest expression percentage level.

CLASP cancer specific expression profile: In order to fulfill the cancer specific criteria, genes must exhibit 0 expression in normal and libraries associated with tumor and disease but not tumor per se. If the gene then exhibited organ-specificity, the gene was selected as a CLASP target for this profile.

CLASP differentially expressed in cancer profile: Expression levels in tumor libraries in each organ and normal libraries (including normal libraries associated with cancer or disease) for all organs were obtained and statistically analyzed. If the gene exhibited 90% of confidence that it is over-expressed in tumor libraries in the target organ than normal libraries for all organs, it was selected as a CLASP target for this profile.

CLASP maximum tissue differential expression profile: CLASP targets were selected based on ratio of expression in tumor libraries compared to expression in normal libraries (including normal libraries associated with tumor or disease) for each organ regardless of whether the gene exhibited organ-specificity. This profile was divided into 2 sub-profiles, since the ratio of expression cannot be obtained if no expression is present in normal libraries(including normal libraries associated with tumor or disease). In this case, the maximum expression percentage of the gene, as calculated by the occurrence of the gene divided by the occurrence of all genes in the target organ, was used. CLASP selects the top 50 targets for each sub-profile.

Accordingly, CLASP allows the identification of highly expressed organ and cancer specific genes based on the gene expression levels in each tissue organ. CLASP scores for a portion of the BSG of this invention are listed below.

DEX0321_1	SEQ ID NO: 1	CLASP5 CLASP3
DEX0321_4	SEQ ID NO: 4	CLASP5 CLASP1
DEX0321_5	SEQ ID NO: 5	CLASP5
DEX0321_6	SEQ ID NO: 6	CLASP5
DEX0321_7	SEQ ID NO: 7	CLASP5 CLASP3
DEX0321_10	SEQ ID NO: 10	CLASP5
DEX0321_11	SEQ ID NO: 11	CLASP5
DEX0321_12	SEQ ID NO: 12	CLASP5 CLASP4
DEX0321_13	SEQ ID NO: 13	CLASP5 CLASP4
DEX0321_14	SEQ ID NO: 14	CLASP5 CLASP3 CLASP4
DEX0321_15	SEQ ID NO: 15	CLASP5 CLASP4
DEX0321_16	SEQ ID NO: 16	CLASP5 CLASP4
DEX0321_17	SEQ ID NO: 17	CLASP5 CLASP4
DEX0321_18	SEQ ID NO: 18	CLASP5 CLASP4
DEX0321_21	SEQ ID NO: 21	CLASP5
DEX0321_22	SEQ ID NO: 22	CLASP5
DEX0321_23	SEQ ID NO: 23	CLASP5
DEX0321_24	SEQ ID NO: 24	CLASP5
DEX0321_25	SEQ ID NO: 25	CLASP5
DEX0321_26	SEQ ID NO: 26	CLASP5 CLASP3
DEX0321_27	SEQ ID NO: 27	CLASP5 CLASP4
DEX0321_28	SEQ ID NO: 28	CLASP5
DEX0321_29	SEQ ID NO: 29	CLASP5
DEX0321_30	SEQ ID NO: 30	CLASP5 CLASP4
DEX0321_31	SEQ ID NO: 31	CLASP5
DEX0321_32	SEQ ID NO: 32	CLASP5
DEX0321_33	SEQ ID NO: 33	CLASP5 CLASP4
DEX0321_34	SEQ ID NO: 34	CLASP5 CLASP4
DEX0321_35	SEQ ID NO: 35	CLASP5
DEX0321_36	SEQ ID NO: 36	CLASP5
DEX0321_37	SEQ ID NO: 37	CLASP5
DEX0321_38	SEQ ID NO: 38	CLASP5 CLASP4 CLASP3
DEX0321_41	SEQ ID NO: 41	CLASP5
DEX0321_42	SEQ ID NO: 42	CLASP5
DEX0321_43	SEQ ID NO: 43	CLASP5 CLASP4 CLASP3
DEX0321_44	SEQ ID NO: 44	CLASP5

DEX0321_45	SEQ ID NO: 45	CLASP2 CLASP1
DEX0321_46	SEQ ID NO: 46	CLASP2 CLASP1
DEX0321_47	SEQ ID NO: 47	CLASP5 CLASP4 CLASP3

In addition the expression values for each organ in the format 9 - 0.9999 are listed. Each box first lists the given organ, then it lists a number representing the percentage of the expression of the gene in the given organ.

321_1	MAM .0085	BRN .0002	ADR .0015	BLV .0016	UTR .0019
321_4	MAM .0028	BRN .0001	FTS .0001	BRN .0002	INL .0004
321_5	MAM .0009				
321_6	MAM .0009				
321_7	MAM .0383	UNC .004	TNS .0054	PIT .0123	PLE .015
321_10		OVR .0051	FAL .0063		
321_11		OVR .0051	FAL .0063		
321_12	MAM .0047	PRO .0006	INL .0006	UTR .0013	ADR .0015
321_13	MAM .0047	PRO .0006	INL .0006	UTR .0013	ADR .0015
321_14	MAM .0727	BLO .008	BLO .008	UNC .008	UNC .008
321_15	MAM .2073	FAL .0126	PIB .0181	PLE .0299	SPC .035
321_16	MAM .2073	FAL .0126	PIB .0181	PLE .0299	SPC .035
321_17	MAM .529	BMR .1609	SAG .1778	SPC .1899	NOS .198
321_18	MAM .529	BMR .1609	SAG .1778	SPC .1899	NOS .198
321_21	MAM .0005				
321_22	MAM .0005				
321_23	MAM .0005				
321_24	MAM .0005				
321_25		PAN .027	BRN .0319	LIV .0435	ADR .0522
321_26	MAM .0453	ESO .0051	BON .0112	INS .0124	CRD .0136
321_27	MAM .1181	NOS .0147	URE .0225	LIV .0246	BON .0394
321_28	MAM .0005				
321_29	MAM .0009	OVR .001			
321_30	MAM .0052	CON .0023	PNS .0023	PAN .0024	ADR .003

321_31	MAM .0142	PRO .013	ADR .0149	LNG .0156	
321_32	MAM .0142	PRO .013	ADR .0149	LNG .0156	
321_33	MAM .0057	FTS .0006	OVR .001	BLD .0016	TST .0027
321_34	MAM .0057	FTS .0006	OVR .001	BLD .0016	TST .0027
321_35	MAM .0213	OVR .0226	BLD .0241	INL .0275	LNG .0374
321_36	MAM .0213	OVR .0226	BLD .0241	INL .0275	LNG .0374
321_37	MAM .0123	PNS .0023	INL .0032	PRO .0034	BON .0056
321_38	MAM .0151	KID .0013	KID .0013	FTS .0015	FTS .0015
321_41	MAM .0024	LIV .0019			
321_42	MAM .0024	LIV .0019			
321_43	MAM .042	SEB .0104	SEB .0104	BON .0169	BON .0169
321_44	MAM .0005				
321_45	MAM .0053	LNG .0003	LMN .0034	TNS .0049	LMN .0099
321_46	MAM .0053	LNG .0003	LMN .0034	TNS .0049	LMN .0099
321_47	MAM .8365	PLE .0449	PLE .0449	SPC .085	SPC .085

Abbreviation for tissues:
ADR Adrenal Glands
BLD Bladder
BLO Blood
BLV Blood Vessels
BMR Bone Marrow
BON Bones
BRN Brain
CON Connective Tissue
CRD Heart
ESO Esophagus
FAL Fallopian Tubes
FTS Fetus
INL Intestine, Large
INS Intestine, Small
KID Kidney
LIV Liver
LMN Lymphoid Tissue
LNG Lung
MAM Breast
NOS Nose
OVR Ovary
PAN Pancreas
PIB Pineal Body
PIT Pituitary Gland
PLE Pleura
PNS Penis
PRO Prostate
SAG Salivary Glands
SEB Seminal Vesicles
SPC Spinal Cord
TNS Tonsil / Adenoids
TST Testis
UNC Mixed Tissues
URE Ureter
UTR Uterus

Based on sequence alignment with the human genome, the following chromosomal locations were assigned. The mapping of the nucleic acid ("NT") SEQ ID NO; NT DEX ID; Parent NT DEX ID; chromosomal location (if known); open reading frame (ORF) location; amino acid ("AA") SEQ ID NO; AA DEX ID; and Parent AA DEX ID are shown in the table below

SEQ ID NO	DEX NT SEQ ID	PARENT DEX NT	Microarray IN	Chromo Map	ORF Loc	SEQ ID NO	DEX AA SEQ ID	PARENT DEX AA
1	DEX0432_001.nt.1	DEX0321_1	mry4259	8				
2	DEX0432_002.nt.1	DEX0321_2	mry4507	6				
3	DEX0432_003.nt.1	DEX0321_3	mry4560	11				
4	DEX0432_004.nt.1	DEX0321_4	mry4787	3				
5	DEX0432_005.nt.1	DEX0321_5	mry4902			95	DEX0432_005.aa.1	DEX0321_48
6	DEX0432_006.nt.1	DEX0321_6	flex mry4902					
7	DEX0432_007.nt.1	DEX0321_7	mry5434	12				
8	DEX0432_008.nt.1	DEX0321_8	mry5572	5				
9	DEX0432_009.nt.1	DEX0321_9	mry5640	5				
10	DEX0432_010.nt.1	DEX0321_10	mry5685	1		96	DEX0432_010.aa.1	DEX0321_49
11	DEX0432_011.nt.1	DEX0321_11	flex mry5685	1				
12	DEX0432_012.nt.1	DEX0321_12	mry5824	1		97	DEX0432_012.aa.1	DEX0321_50
13	DEX0432_013.nt.1	DEX0321_13	flex mry5824	1		98	DEX0432_013.aa.1	DEX0321_51
14	DEX0432_014.nt.1	DEX0321_14	mry5904	2				
15	DEX0432_015.nt.1	DEX0321_15	mry5988	15		99	DEX0432_015.aa.1	DEX0321_52
16	DEX0432_016.nt.1	DEX0321_16	flex mry5988	15		100	DEX0432_016.aa.1	DEX0321_53
17	DEX0432_017.nt.1	DEX0321_17	mry6191	X		101	DEX0432_017.aa.1	DEX0321_54
18	DEX0432_018.nt.1	DEX0321_18	flex mry6191	6		102	DEX0432_018.aa.1	DEX0321_55
19	DEX0432_019.nt.1	DEX0321_19	mry6723	5		103	DEX0432_019.aa.1	DEX0321_56
20	DEX0432_020.nt.1	DEX0321_20	flex mry6723	5				
21	DEX0432_021.nt.1	DEX0321_21	mry6804	2		104	DEX0432_021.aa.1	DEX0321_57
22	DEX0432_022.nt.1	DEX0321_22	flex mry6804	2				
23	DEX0432_023.nt.1	DEX0321_23	mry7407	8		105	DEX0432_023.aa.1	DEX0321_58
24	DEX0432_024.nt.1	DEX0321_24	flex mry7407	8				

25	DEX0432_025.nt.1	DEX0321_25	mry7505	20		106	DEX0432_025.aa.1	DEX0321_59
26	DEX0432_026.nt.1	DEX0321_26	flex mry7505	20		107	DEX0432_026.aa.1	DEX0321_60
27	DEX0432_027.nt.1	DEX0321_27	mry7575	19				
28	DEX0432_028.nt.1	DEX0321_28	mry7689	18				
29	DEX0432_029.nt.1	DEX0321_29	mry7812	2				
30	DEX0432_030.nt.1	DEX0321_30	mry7951	8				
31	DEX0432_031.nt.1	DEX0321_31	mry8181	7		108	DEX0432_031.aa.1	DEX0321_61
32	DEX0432_032.nt.1	DEX0321_32	flex mry8181	7				
33	DEX0432_033.nt.1	DEX0321_33	mry8214	18		109	DEX0432_033.aa.1	DEX0321_62
34	DEX0432_034.nt.1	DEX0321_34	flex mry8214	18				
35	DEX0432_035.nt.1	DEX0321_35	mry8268	1		110	DEX0432_035.aa.1	DEX0321_63
36	DEX0432_036.nt.1	DEX0321_36	flex mry8268	1				
37	DEX0432_036.nt.2			*	73-3844	111	DEX0432_036.aa.2	
38	DEX0432_036.nt.3			*	73-2255	112	DEX0432_036.aa.3	
39	DEX0432_036.nt.4			*	73-1648	113	DEX0432_036.aa.4	
40	DEX0432_036.nt.5			*	73-1621	114	DEX0432_036.aa.5	
41	DEX0432_036.nt.6			*	-	115	DEX0432_036.aa.6	
42	DEX0432_036.nt.7			*	72-	116	DEX0432_036.aa.7	
43	DEX0432_036.nt.8			1q22	184-1612	117	DEX0432_036.aa.8	
44	DEX0432_036.nt.9			1q22	184-1612	117	DEX0432_036.aa.8	
45	DEX0432_036.nt.10			1q22	836-1530	118	DEX0432_036.aa.10	
46	DEX0432_036.nt.11			1q22	195-519	119	DEX0432_036.aa.11	
47	DEX0432_036.nt.12			1q22	184-1639	120	DEX0432_036.aa.12	

48	DEX0432_036.nt.13			1q22	184-1576	121	DEX0432_036.aa.13	
49	DEX0432_036.nt.14			1q22	49-1625	122	DEX0432_036.aa.14	
50	DEX0432_036.nt.15			1q22	184-1492	123	DEX0432_036.aa.15	
51	DEX0432_036.nt.16			1q22	184-1006	124	DEX0432_036.aa.16	
52	DEX0432_036.nt.17			1q22	184-1237	125	DEX0432_036.aa.17	
53	DEX0432_036.nt.18			1q22	184-1639	120	DEX0432_036.aa.12	
54	DEX0432_036.nt.19			1q22	184-1135	126	DEX0432_036.aa.19	
55	DEX0432_036.nt.20			1q22	184-877	127	DEX0432_036.aa.20	
56	DEX0432_036.nt.21			1q22	184-2029	128	DEX0432_036.aa.21	
57	DEX0432_036.nt.22			1q22	184-1303	129	DEX0432_036.aa.22	
58	DEX0432_036.nt.23			1q22	184-955	130	DEX0432_036.aa.23	
59	DEX0432_036.nt.24			1q22	184-1663	131	DEX0432_036.aa.24	
60	DEX0432_036.nt.25			1q22	184-1634	132	DEX0432_036.aa.25	
61	DEX0432_036.nt.26			1q22	184-637	133	DEX0432_036.aa.26	
62	DEX0432_036.nt.27			1q22	184-691	134	DEX0432_036.aa.27	
63	DEX0432_036.nt.28			1q22	1-242	135	DEX0432_036.aa.28	
64	DEX0432_036.nt.29			1q22	184-1378	136	DEX0432_036.aa.29	
65	DEX0432_036.nt.30			1q22	1-113	137	DEX0432_036.aa.30	

66	DEX0432_036.nt.31			1q22	184-979	138	DEX0432_036.aa.31	
67	DEX0432_036.nt.32			1q22	184-910	139	DEX0432_036.aa.32	
68	DEX0432_036.nt.33			1q22	34-313	140	DEX0432_036.aa.33	
69	DEX0432_036.nt.34			1q22	380-1643	141	DEX0432_036.aa.34	
70	DEX0432_036.nt.35			1q22	39-1497	142	DEX0432_036.aa.35	
71	DEX0432_036.nt.36			1q22	184-952	143	DEX0432_036.aa.36	
72	DEX0432_036.nt.37			1q22	184-1738	144	DEX0432_036.aa.37	
73	DEX0432_036.nt.38			1q22	417-960	145	DEX0432_036.aa.38	
74	DEX0432_036.nt.39			1q22	184-883	146	DEX0432_036.aa.39	
75	DEX0432_036.nt.40			1q22	184-910	139	DEX0432_036.aa.32	
76	DEX0432_036.nt.41			1q22	184-1375	147	DEX0432_036.aa.41	
77	DEX0432_036.nt.42			1q22	184-1162	148	DEX0432_036.aa.42	
78	DEX0432_036.nt.43			1q22	184-1414	149	DEX0432_036.aa.43	
79	DEX0432_036.nt.44			1q22	184-1324	150	DEX0432_036.aa.44	
80	DEX0432_036.nt.45			1q22	1-334	151	DEX0432_036.aa.45	
81	DEX0432_036.nt.46			1q22	184-568	152	DEX0432_036.aa.46	
82	DEX0432_036.nt.47			1q22	716-1040	119	DEX0432_036.aa.11	
83	DEX0432_036.nt.48			1q22	-	153	DEX0432_036.aa.48	
84	DEX0432_037.nt.1	DEX0321_37	mry8376	17				
85	DEX0432_038.nt.1	DEX0321_38	mry8420	1				

86	DEX0432_039.nt.1	DEX0321_39	mry8476	9		154	DEX0432_039.aa.1	DEX0321_64
87	DEX0432_040.nt.1	DEX0321_40	flex mry8476	9				
88	DEX0432_041.nt.1	DEX0321_41	mry8502	17		155	DEX0432_041.aa.1	DEX0321_65
89	DEX0432_042.nt.1	DEX0321_42	flex mry8502	17				
90	DEX0432_043.nt.1	DEX0321_43	mry8644	1				
91	DEX0432_044.nt.1	DEX0321_44	mry8764	18				
92	DEX0432_045.nt.1	DEX0321_45	mry8936	4		156	DEX0432_045.aa.1	DEX0321_66
93	DEX0432_046.nt.1	DEX0321_46	flex mry8936	4				
94	DEX0432_047.nt.1	DEX0321_47	mry9072	2				

The microarray sequence identifications, extended sequences based human genome (flex) and predicted peptide sequences for each of the targets are listed below:

SEQ ID NO	Microarray IN	Predicted Peptide
DEX0321_1	mry4259	
DEX0321_2	mry4507	
DEX0321_3	mry4560	
DEX0321_4	mry4787	
DEX0321_5	mry4902	DEX0321_48
DEX0321_6	flex mry4902	
DEX0321_7	mry5434	
DEX0321_8	mry5572	
DEX0321_9	mry5640	
DEX0321_10	mry5685	DEX0321_49
DEX0321_11	flex mry5685	
DEX0321_12	mry5824	DEX0321_50
DEX0321_13	flex mry5824	DEX0321_51
DEX0321_14	mry5904	
DEX0321_15	mry5988	DEX0321_52
DEX0321_16	flex mry5988	DEX0321_53
DEX0321_17	mry6191	DEX0321_54
DEX0321_18	flex mry6191	DEX0321_55
DEX0321_19	mry6723	DEX0321_56
DEX0321_20	flex mry6723	
DEX0321_21	mry6804	DEX0321_57
DEX0321_22	flex mry6804	
DEX0321_23	mry7407	DEX0321_58
DEX0321_24	flex mry7407	
DEX0321_25	mry7505	DEX0321_59
DEX0321_26	flex mry7505	DEX0321_60
DEX0321_27	mry7575	
DEX0321_28	mry7689	
DEX0321_29	mry7812	
DEX0321_30	mry7951	
DEX0321_31	mry8181	DEX0321_61
DEX0321_32	flex mry8181	
DEX0321_33	mry8214	DEX0321_62
DEX0321_34	flex mry8214	
DEX0321_35	mry8268	DEX0321_63
DEX0321_36	flex mry8268	
DEX0321_37	mry8376	
DEX0321_38	mry8420	
DEX0321_39	mry8476	DEX0321_64
DEX0321_40	flex mry8476	
DEX0321_41	mry8502	DEX0321_65
DEX0321_42	flex mry8502	
DEX0321_43	mry8644	
DEX0321_44	mry8764	
DEX0321_45	mry8936	DEX0321_66
DEX0321_46	flex mry8936	
DEX0321_47	mry9072	

Example 1A: Suppression subtractive hybridization (Clontech PCR-SELECT)

Clontech PCR-SELECT is a PCR based subtractive hybridization method designed to selectively enrich for cDNAs corresponding to mRNAs differentially expressed between two mRNA populations (Diatchenko et al, Proc. Natl. Acad. Sci. USA, Vol. 93,

pp. 6025-6030, 1996). Clontech PCR-SELECT is a method for enrichment of differentially expressed mRNAs based on a selective amplification. cDNA is prepared from the two mRNA populations which are to be compared (Tester: cDNA population in which the differentially expressed messages are sought and Driver: cDNA population in which the differentially expressed transcripts are absent or low). The tester sample is separated in two parts and different PCR adapters are ligated to the 5' ends. Each tester is separately annealed to excess driver (first annealing) and then pooled and again annealed (second annealing) to excess driver. During the first annealing sequences common to both populations anneal. Additionally the concentration of high and low abundance messages are normalized since annealing is faster for abundant molecules due to the second order kinetics of hybridization. During the second annealing cDNAs unique or overabundant to the tester can anneal together. Such molecules have different adapters at their ends. The addition of additional driver during the second annealing enhances the enrichment of the desired differentially expressed sequences. During subsequent PCR, molecules that have different adapters at each end amplify exponentially. Molecules which have identical adapters, or adapters at only one end, or no adapters (driver sequences) either do not amplify or undergo linear amplification. The end result is enrichment for cDNAs corresponding to differentially expressed messages (unique to the tester or upregulated in the tester). This technique was used to identify transcripts unique to breast tissue or messages overexpressed in breast cancer. Pairs of matched samples isolated from the same patient, a cancer sample, and the "normal" adjacent tissue from the same tissue type were utilized. The mRNA from the cancer tissue is used as the "tester", and the non-cancer mRNA as a "driver". The non-cancer "driver" is from the same individual and tissue as the cancer sample (Matched). Alternatively, the "driver" can be from a different individual but the same tissue as the tumor sample (unmatched). In some cases mixtures of mRNAs derived from non-cancer tissues types different from the cancer tissue type are also used as "drivers". The last approach allows the identification of transcripts whose expression is specific or upregulated in the cancer tissue type analyzed. Such transcripts may or may not be cancer specific in their expression.

Several subtracted libraries were generated for breast tissue. The product of the subtraction experiments was used to generate cDNA libraries. These cDNA libraries contain Expressed Sequence Tags (ESTs) from genes that are breast cancer specific, or upregulated in breast tissue. Randomized clones picked from each cDNA PCR Select

library were sequenced and the genes identified by a systematic analysis of the sequence data against the LIFESEQ Gold database available from Incyte Pharmaceuticals, Palo Alto. All of the lead sequences were discoved using subtractions.

Example 1b: Alternative Splice Variants

We identified gene transcripts associated with cancer disease, development, or progression using cloning experiments, the Gencarta™ tools software (Compugen Ltd., Tel Aviv, Israel), and a variety of public and proprietary databases. These transcripts are either novel splice variant sequences which differ from a previously defined sequence or new uses of known sequences. In general the previously defined sequence for a transcript family is annotated as DEX0432_XXX.nt.1 and the novel variants are annotated as DEX0432_XXX.nt.2, DEX0432_XXX.nt.3, etc. The novel variant DNA sequences encode novel proteins which differ from a previously defined protein sequence. In relation to the nucleotide sequence naming convention, the previously defined amino acid sequence is annotated DEX0432_XXX.aa.1 and the novel variants annotated as DEX0432_XXX.aa.2, etc.

EST Support

The alternative splice variants are predicted by computational analysis of Expressed Sequence Tags (ESTs) derived form public and proprietary cDNA libraries and genomic information. A novel transcript may be supported by numerous ESTs.

SAGE Support

Serial Analysis of Gene Expression (SAGE) tag data analysis is preformed on the novel splice variants. Gencarta™ tools (Compugen Ltd., Tel Aviv, Israel) report SAGE tag data for individual transcripts when available. SAGE data includes the SAGE tag sequence for the novel transcripts, expression level (as a ratio) of the SAGE tag in tissue samples, the source or tissue, state or disease condition of the tissue, tissue sample type, and a description of the tissue samples. SAGE tag data analysis results are disclosed and discussed in each transcript section below.

Sequence Alignment Support

Alignments of previously identified and novel splice variant sequences are performed to confirm unique portions of splice variant nucleic acid and amino acid sequences. The alignments are done using the Needle program in the European Molecular

Biology Open Software Suite (EMBOSS) version 2.2.0 available at www.emboss.org from EMBnet (http://www.embnet.org). Default settings are used unless otherwise noted. The Needle program in EMBOSS implements the Needleman-Wunsch algorithm. Needleman, S. B., Wunsch, C. D., *J. Mol. Biol.* 48:443-453 (1970).

It is well know to those skilled in the art that implication of alignment algorithms by various programs may result in minor changes in the generated output. These changes include but are not limited to: alignment scores (percent identity, similarity, and gap), display of nonaligned flanking sequence regions, and number assignment to residues. These minor changes in the output of an alignment do not alter the physical characteristics of the sequences or the differences between the sequences, e.g. regions of homology, insertions, or deletions. Descriptions of alignments are provided in each splice variant family section.

<u>DEX0432_035.nt.1, DEX0432_036.nt.1 (Mam096); Splice Variants DEX0432_036.nt.2 – DEX0432_036.nt.48 (Mam096v)</u>

Novel transcripts of the Mam096 family which include variants DEX0432_35.nt.1 and DEX0432_36.nt.1 – DEX0432_36.nt.48, were discovered using the methods described above. The use of "Mam096" herein refers to the transcript family and is meant to include the variants known in the literature. Mam096 has also been identified as Glycoprotein 39 3' fragment in JP 07051065-A; Human cancer associated gene sequence SEQ ID NO:19 in WO 005/5350-A1; Thyroid cancer related gene sequence SEQ ID NO:5876 in WO 01/94629-A2; and Human gene expression profile polynucleotide SEQ ID NO 339 in WO 02/74979-A2 which are herein incorporated by reference.

In addition to the nomenclature from the patents above, there are many synonyms for Mam096 in the literature. They include Mucin 1 precursor (MUC-1), Polymorphic epithelial mucin (PEM) (PEMT), Episialin, Tumor-associated mucin, Carcinoma-associated mucin, Tumor-associated epithelial membrane antigen (EMA), H23AG, Peanut- reactive urinary mucin (PUM), Breast carcinoma-associated antigen DF3, and CD227 antigen. Lan,M.S., *et al.* (1990) *J. Biol. Chem.* 265:15294-15299; Ligtenberg,M.J.L., *et al.* (1990) *J. Biol. Chem.* 265:5573-5578; Gendler,S.J., *et al.* (1990) *J. Biol. Chem.* 265:15286-15293; Lancaster,C.A., *et al.* (1990) *Biochem. Biophys. Res. Commun.* 173:1019-1029; Wreschner,D.H., *et al.* (1990) *Eur. J. Biochem.* 189:463-473; Hareuveni,M., *et al.* (1990) *Eur. J. Biochem.* 189:475-486; Tsarfaty,I. *Et al.*(1990) *Gene*

93:313-318; Zrihan-Licht,S., *et al.* (1994) *Eur. J. Biochem.* 224:787-795; Oosterkamp,H.M. *et al.* (1997) *Int. J. Cancer* 72:87-94; Zhang,L.X. *et al.* Molecular cloning of an isoform of MUC1, MUC1/Y.Submitted FEB-1999 to the EMBL GenBank DDBJ databases (Isoform 7); Zhang,L.X. *et al.* Cloning of a new potential secreted short variant form of MUC1 mucin in epithelial cancer cell line.Submitted FEB-2001 to the EMBL GenBank DDBJ databases (Isoform 9); Gendler,S.J. *et al.* (1988) *J. Biol. Chem.* 263:12820-12823; Abe,M. *et al.* (1989) *Biochem. Biophys. Res. Commun.* 165:644-649; Weiss,M. *et al.* (1996) *Int. J. Cancer* 66:55-59; Yu,C.J. *et al.* (1996) *Oncology* 53:118-126; Buluwela,L. *et al.* Submitted OCT-1992 to the EMBL GenBank DDBJ databases (ISOFORMS 3 AND 4); Mueller,S. *et al.* (1997) *J. Biol. Chem.* 272:24780-24793; Mueller,S. *et al.* (1999) *J. Biol. Chem.* 274:18165-18172; Engelmann,K. *et al.* (2001) *J. Biol. Chem.* 276:27764-27769; Baruch,A. *et al.* (1999) *Cancer Res.* 59:1552-1561); Parry,S. *et al.* (2001) *Biochem. Biophys. Res. Commun.* 283:715-720; Wreschner,D.H., *et al.* (2002) *Protein Sci.* 11:698-706; and Zrihan-Licht,S., *et al.* (1994) *FEBS Lett.* 356:130-136.

Mucin-1 (MUC1) is a type I membrane protein and contains 1 SEA domain. Two known secreted forms (5 and 9) are also produced. Mucin-1 may play a role in adhesive functions and in cell-cell interactions, metastasis and signaling. It may also provide a protective layer on epithelial surfaces. Direct or indirect interaction with actin cytoskeleton. Isoform 7 behaves as a receptor and binds the secreted isoform 5. The binding induces the phosphorylation of the isoform 7, alters cellular morphology and initiates cell signaling. Additionally, Isoform 7can bind to grb2 adapter protein. The cleaved form of Mucin-1 (isoform 1) forms a tight heterodimer with the released C-terminal peptide (which is first secreted to be extracellular). MUC1 is expressed on ductal epithelial cells and on activated T-cells. Aberrantly glycosylated forms are expressed in human epithelial tumors, such as breast or ovarian cancer and also in non-epithelial tumor cells. Isoform 7 is expressed only in tumoral cells.

MUC1 is highly glycosylated (N-and O-linked carbohydrates and sialic acid). In the 20 amino acid tandem repeat positions 5 (ser), 6 (thr), 14 (thr), 15 (ser) and 19 (thr) are O-glycosylated (galnac). The average density of O-glycosylated sites within repeat peptides varies with cell differentiation from about 50% in lactation-associated MUC1 to over 90% in a variety of breast cancer cells. Isoforms 1 and 7 undergo transphosphorylation on serine and tyrosine residues.

The MUC1 number of repeats is highly polymorphic. It varies from 21 to 125 in the northern european population. The most frequent alleles contains 41 and 85 repeats. The tandemly repeated icosapeptide underlies polymorphism at three positions: PAPGSTAP[PAQT]AHGVTSAP[DT/ES]R, DT -> ES and the single replacements P -> A, P -> Q and P-> T. The most frequent replacement DT > ES occurs in up to 50% of the repeats. SWISS-PROT accession Numbers: P15941, P13931, P15942, P17626, Q14128, Q14876, Q16437, Q16442, Q16615, Q9BXA4, Q9UE75, Q9UE76, Q9UQL1, Q9Y4J2. SWISS-PROT is accessible at http://www.ebi.ac.uk/swissprot/ .

Splice Variant Nucleotides

Novel splice variants have been identified for the Mam096 family, DEX0432_035.nt.1 and DEX0432_036.nt.1 – DEX0432_036.nt.48. These novel transcripts are located in the same genomic region as MUC1 family. Mam096 variants contain novel exon additions and deletions which encode for unique amino acid sequences. These unique amino acid sequence provide new proteins to be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer. The unique nucleotide sequences in these new transcript can be used as nucleic acid probes for the diagnosis and/or treatment of cancer.

Alignments of the DNA sequences for Mam096 family display regions of similarity and difference between transcripts.

Splice Variant Polypeptides

The nucleotide sequences of the novel splice variants for Mam096 (DEX0432_035.nt.1 and DEX0432_036.nt.1 – DEX0432_036.nt.48), encode novel amino acid sequences DEX0432_35.aa.1 and DEX0432_36.aa.2 – DEX0432_36.aa.46. The novel amino acid sequences are novel Mam096 protein variants. These proteins contain novel features including unique epitopes, new cellular locilazations, and altered function. Novel features can of the proteins can be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer.

Alignments of the amino acid sequences for Mam096 family display regions of similarity and difference between transcripts.

Altogether, splice variant sequence analysis, EST support, and SAGE tag data are indicative of SEQ ID NO: 1-94 and SEQ ID NO: 95-156 being a diagnostic marker and/or a therapeutic target for cancer.

Example 1c: RT-PCR Analysis

To detect the presence and tissue distribution of a particular splice variant Reverse Transcription-Polymerase Chain Reaction (RT-PCR) is performed using cDNA generated from a panel of tissue RNAs. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and; Kawasaki ES *et al.*, *PNAS* 85(15):5698 (1988). Total RNA is extracted from a variety of tissues and first strand cDNA is prepared with reverse transcriptase (RT). Each panel includes 23 cDNAs from five cancer types (lung, ovary, breast, colon, and prostate) and normal samples of testis, placenta and fetal brain. Each cancer set is composed of three cancer cDNAs from different donors and one normal pooled sample. Using a standard enzyme kit from BD Bioscience Clontech (Mountain View, CA), the target transcript is detected with sequence-specific primers designed to only amplify the particular splice variant. The PCR reaction is run on the GeneAmp PCR system 9900 (Applied Biosystem, Foster City, CA) thermocycler under optimal conditions. One of ordinary skill can design appropriate primers and determine optimal conditions. The amplified product is resolved on an agarose gel to detect a band of equivalent size to the predicted RT-PCR product. A band indicates the presence of the splice variant in a sample. The relation of the amplified product to the splice variant is subsequently confirmed by DNA sequencing.

The RT-PCR experiments confirm the physical existence of SEQ ID NO: 1-94 in a biological sample. RT-PCR experiments results include cancer tissue(s) detected in, predicted band length, and experimentally confirmed band length for each transcript.

RT-PCR results confirm the presence SEQ ID NO: 1-94 in biologic samples and distinguish between related transcripts.

Example 1d: Secretion Assay

To determine if a protein encoded by a novel splice variant is secreted from cells a secretion assay is preformed. pcDNA3.1 clones containing transcripts from the same family encoding different forms of proteins are transfected into 293T cells using the Superfect transfection reagent (Qiagen, Valencia CA). Transfected cells are incubated for 28 hours before the media is collected and immediately spun down to remove any detached cells. The adherent cells are solubilized with lysis buffer (1% NP40, 10mM sodium phosphate pH7.0, and 0.15M NaCl). The lysed cells are collected and spun down and the supernatant extracted as cell lysate. Western immunoblot is carried out in the

following manner: 15ul of the cell lysate and media are run on 4-12% NuPage Bis-Tris gel (Invitrogen, Carlsbad CA), and blotted onto a PVDF membrane (Invitrogen, Carlsbad CA). The blot is incubated with polyclonal anti-TRAILR2 primary antibody (Imgenex, San Diego CA) and polyclonal goat anti-rabbit-peroxidase secondary antibody (Sigma-Aldrich, St. Louis MO). The blot is developed with the ECL Plus chemiluminescent detection reagent (Amersham BioSciences, Piscataway NJ).

Secretion assay results are indicative of SEQ ID NO: 95-156 being a diagnostic marker and/or therapeutic target for cancer.

Example 2A: Custom Microarray Experiment—Breast Cancer

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the in-situ synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature Biotechnology 19:342-347). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two differents 60mers were designed for each gene of interest. All microarray experiments were two-color experiments and were preformed using Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was polyA+ RNA isolated from cancer tissue from a single individual and the reference sample was a pool of polyA+ RNA isolated from normal tissues of the same organ as the cancerous tissue (i.e. normal breast tissue in experiments with breast cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent in-situ hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon). A total of 29 experiments comparing the expression patterns of breast cancer derived polyA+ RNA (15 squamous cell carcinomas, 14 adenocarcinomas) to polyA+ RNA isolated from a pool of 12 normal breast tissues were analyzed.

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression

analysis was performed using only experiments that meet certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control + 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Arrays with poor detection limits were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employ user defined parameters to identify quality data. Only those features that meet the threshhold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background + 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors and saturated features were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up- and down- regulated genes were identified using criteria for percentage of valid values obtained, and the percentage of experiments in which the gene is up- or down-regulated. These criteria were set independently for each data set, depending on the size and the nature of the data set. The results for the significantly upregulated and downregulated genes are shown in Table 1. The first three columns of the table contain information about the sequence itself (Oligo ID, Parent ID, and Patent#), the next 3 columns show the results obtained. '%valid' indicates the percentage of 29 unique experiments total in which a valid expression value was obtained, '%up' indicates the percentage of 29 experiments in which up-regulation of at least 2.5-fold was observed, and '%down' indicates the percentage of the 29 experiments in which down-regulation of at least 2.5-fold was observed. The last column in Table 1 describes the location of the microarray probe (oligo) relative to the parent sequence for upregulated genes. Table 2 describes the results and oligo locations for down-regulated genes. Tables 3 and 4 describe the results and oligo locations for up and down regulated genes, respectively.

Table 1. Sensitivity data for up-regulated genes. Data reported for Parent IDs (Par. ID) denoted by * are calculated based on lower-voltage PMT scans due to saturation in higher-voltage PMT scans (extremely high expression levels).

DEX ID	Par. ID	% valid n=35	% up n=35	% up ST1 n=9	% up ST2, 3 n=26	Oligo ID	Start Pos. Par. Seq	Stop Pos. Par. Seq	Start Pos. FLEXS	Stop Pos. FLEX S
DEX0321_4	4787	27.8	11.1	33.3	3.7	11104	1497	1556		
DEX0321_4	4787	52.8	8.3	33.3	0	11105	1278	1337		
DEX0321_8	5572	100	38.9	33.3	40.7	21873	133	192		
DEX0321_14	5904	94.4	75	77.8	74.1	37806	18	77		
DEX0321_23	7407	86.1	30.6	11.1	37	26346	658	717	782	841
DEX0321_23	7407	63.9	2.8	0	3.7	26347	583	642	707	766
DEX0321_28	7689	77.8	50	55.6	48.1	12645	298	357		
DEX0321_28	7689	88.9	47.2	55.6	44.4	12646	270	329		
DEX0321_30	7951	97.2	8.3	11.1	7.4	15120	194	253		
DEX0321_30	7951	97.2	25	33.3	22.2	15121	101	160		
DEX0321_31	8181	100	41.7	22.2	48.1	17232	585	644	997	1056
DEX0321_31	8181	97.2	41.7	33.3	44.4	17233	371	430	783	842
DEX0321_33	8214	100	33.3	22.2	37	17942	178	237	446	505
DEX0321_33	8214	100	30.6	33.3	29.6	17943	120	179	388	447
DEX0321_35	8268	100	36.1	66.7	25.9	18206	571	630	572	631
DEX0321_35	8268	97.2	19.4	33.3	14.8	18207	211	270	211	270
DEX0321_38	8420	100	41.7	55.6	37	20235	973	1032		
DEX0321_38	8420	100	41.7	55.6	37	20236	707	766		
DEX0321_43	8644	94.4	36.1	33.3	37	26841	1431	1490		
DEX0321_44	8764	94.4	33.3	55.6	25.9	28227	1154	1213		
DEX0321_44	8764	58.3	11.1	22.2	7.4	28228	1056	1115		

Table 2. Sensitivity data for down-regulated genes. Data reported for Parent IDs denoted by * are calculated based on lower-voltage PMT scans due to saturation in higher-voltage PMT scans (extremely high expression levels).

DEX ID	Par. ID	% valid n=35	% dn n=35	% dn ST1 n=9	% dn ST2, 3 n=26	Oligo ID	Start Pos. Par. Seq	Stop Pos. Par. Seq	Start Pos. FLEXS	Stop Pos. FLEXS
DEX0321_1	4259	100	30.6	33.3	29.6	20385	1398	1457		
DEX0321_1	4259	100	16.7	0	22.2	20386	1325	1384		
DEX0321_2	4507	100	33.3	44.4	29.6	17992	166	225		
DEX0321_2	4507	100	33.3	44.4	29.6	17993	146	205		
DEX0321_3	4560	100	33.3	55.6	25.9	23876	102	161		
DEX0321_5	4902	75	25	44.4	18.5	23856	789	848	789	848
DEX0321_5	4902	91.7	0	0	0	23857	646	705	646	705
DEX0321_7	5434	97.2	33.3	44.4	29.6	34823	289	348		
DEX0321_9	5640	86.1	36.1	22.2	40.7	41765	209	268		
DEX0321_10	5685	97.2	63.9	66.7	63	21328	160	219	692	751
DEX0321_12	5824	100	47.2	22.2	55.6	15681	152	211	2721	2780
DEX0321_12	5824	100	22.2	22.2	22.2	15682	126	185	2695	2754
DEX0321_15	5988	69.4	30.6	44.4	25.9	26124	347	406	654	713
DEX0321_15	5988	5.6	2.8	0	3.7	26125	306	365	613	672
DEX0321_17	6191	80.6	16.7	33.3	11.1	30141	121	180	283	342
DEX0321_19	6723	86.1	66.7	66.7	66.7	18882	654	713	654	713
DEX0321_19	6723	61.1	55.6	22.2	66.7	18883	610	669	610	669
DEX0321_21	6804	83.3	25	33.3	22.2	38921	582	641	633	692
DEX0321_21	6804	69.4	19.4	22.2	18.5	38922	451	510	502	561
DEX0321_25	7505*	100	36.1	22.2	38.5	15020	143	202		
DEX0321_25	7505*	100	33.3	22.2	38.5	15021	97	156		
DEX0321_27	7575	72.2	8.3	0	11.1	40479	76	135		
DEX0321_27	7575	100	69.4	55.6	74.1	40480	11	70		
DEX0321_29	7812	97.2	75	66.7	77.8	13936	144	203		
DEX0321_37	8376	94.4	0	0	0	19390	2438	2497		
DEX0321_37	8376	100	38.9	44.4	37	19391	2151	2210		
DEX0321	8476	97.2	2.8	11.1	0	20533	688	747	802	861

_39										
DEX0321_39	8476	97.2	38.9	66.7	29.6	20534	590	649	704	763
DEX0321_41	8502	97.2	27.8	11.1	33.3	20707	937	996	1749	1808
DEX0321_41	8502	97.2	30.6	22.2	33.3	20708	897	956	1709	1768
DEX0321_45	8936	88.9	63.9	55.6	66.7	31658	377	436	463	522
DEX0321_45	8936	100	38.9	33.3	40.7	31659	256	315	342	401
DEX0321_47	9072	94.4	11.1	33.3	3.7	33550	1310	1369		
DEX0321_47	9072	100	30.6	44.4	25.9	33551	742	801		

Table 3. Sensitivity data for stage-specific up-regulated genes.

DEX ID	Parent ID	%valid n=35	% up n=35	% up ST1 n=9	% up ST2,3 n=26	OligoID	Start Pos. Par. Seq	Stop Pos. Par. Seq	Start Pos. FLEXS	Stop Pos. FLEXS
DEX0321_4	4787	27.8	11.1	33.3	3.7	11104	1497	1556		
DEX0321_4	4787	52.8	8.3	33.3	0	11105	1278	1337		
DEX0321_23	7407	86.1	30.6	11.1	37	26346	658	717	782	841
DEX0321_23	7407	63.9	2.8	0	3.7	26347	583	642	707	766
DEX0321_35	8268	100	36.1	66.7	25.9	18206	571	630	572	631
DEX0321_35	8268	97.2	19.4	33.3	14.8	18207	211	270	211	270
DEX0321_44	8764	94.4	33.3	55.6	25.9	28227	1154	1213		
DEX0321_44	8764	58.3	11.1	22.2	7.4	28228	1056	1115		

Table 4. Sensitivity data for stage-specific down-regulated genes. . Data reported for Parent IDs denoted by * are calculated based on lower-voltage PMT scans due to saturation in higher-voltage PMT scans (extremely high expression levels).

DEX ID	Parent ID	%valid n=35	% dn n=35	% dn ST1 n=9	% dn ST2,3 n=26	OligoID	Start Pos. Par. Seq	Stop Pos. Par. Seq	Start Pos. FLEXS	Stop Pos. FLEXS
DEX0321_3	4560	100	33.3	55.6	25.9	23876	102	161		
DEX0321_5	4902	75	25	44.4	18.5	23856	789	848	789	848
DEX0321_5	4902	91.7	0	0	0	23857	646	705	646	705
DEX0321_12	5824	100	47.2	22.2	55.6	15681	152	211	2721	2780
DEX0321_12	5824	100	22.2	22.2	22.2	15682	126	185	2695	2754

Example 2B: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman® probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman®) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity

of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the comparative method (User Bulletin #2: ABI PRISM 7900 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman® probes specific to each target gene. The results are analyzed using the ABI PRISM 7900 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the BSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to the calibrator. Normal RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the BSNA in pairs of matched samples may also be determined. A matched pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. All the values are compared to the calibrator.

In the analysis of matching samples, the BSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples. Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer state (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matched samples tested are indicative of SEQ ID NO: XX(3) and the encoded protein SEQ ID NO: YY(3) being a diagnostic marker for cancer.

Mam097 (DEX0432_028.nt.1)

The relative expression level of Mam097 in various tissue samples is included below. Tissue samples include 77 pairs of matching samples, 8 non matched cancer samples, and 36 normal samples, all from various tissues annotated in the table. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Of the normal samples 6 were blood samples which measured the expression level of the BSNA in blood cells. Additionally, 2 prostatitis, and 4 Benign Prostatic Hyperplasia (BPH) samples are included. All the values are compared to kidney normal sample KID55KD (calibrator).

The table below contains the relative expression level values for the sample as compared to the calibrator. The table includes the Sample Name, Tissue type, and expression lelvel values for the following samples: Cancer (CAN), Normal Adjacent Tissue (NAT), Normal Tissue (NRM), Benign Prostatic Hyperplasia (BPH), and Prostatitis (PROST).

Sample	Tissue	CAN	NAT	NRM	BPH	PROST
mam522	Breast	689.05				
MamS854	Breast	264.37	228.55			
mamS516	Breast	1292.87	15.09			
mamS621	Breast	2741.5				
MamS570	Breast	0.17	86.76			
MamB011	Breast	17.7	155.21			
Mam19DN	Breast	115.71	115.97			
Mam781M	Breast	75.37	20.36			
MamS699	Breast	205.04	224.93			
Mam543M	Breast	8.17	5.58			
Mam976M	Breast	169.28	5.21			
MamS997	Breast	17.76	9.32			
mam355	Breast	274.93	20.02			
Mam42DN	Breast	236.57	51.85			
Mam76DN	Breast	1106.1	2.1			
Mam01MA	Breast			48.59		
Adr48AD	Adrenal			4.43		
Bld46XK	Bladder	24.18	24.72			
BldTR147	Bladder	14.44	8.12			
Bld520B	Bladder	110.07	45.07			
Bld23BL	Bladder			1.36		
BloB1	Blood			319.96		
BloB3	Blood			222.77		
BloB5	Blood			84.36		

BloB6	Blood			62.79		
BloB11	Blood			100.2		
BloB14	Blood			105.73		
Brn10BR	Brain			3.9		
CvxKS52	Cervix	1.37	12.79			
CvxKS83	Cervix	1.22	8.08			
CvxNKS18	Cervix	0.15	2.16			
CvxNKS81	Cervix	2.47	4.93			
CvxNKS54	Cervix	2.26	3.5			
Cvx1ACB	Cervix			0		
ClnAS43	Colon	0.81	3.98			
ClnAS98	Colon	1.49	1.82			
ClnRS53	Colon	0.84	4.82			
ClnRC01	Colon	6.62	2.98			
ClnSG27	Colon	0.8	11.92			
ClnDC19	Colon	8.99	22.93			
ClnDC63	Colon	6.28	29.18			
ClnCM12	Colon	6.07	8.07			
ClnTX01	Colon	2.7	9.37			
Cln01CL	Colon					
Endo10479	Endometrium	13.55	10.4			
Endo28XA	Endometrium	12.24	39.82			
Endo3AX	Endometrium	33.37	23.05			
Eso1ES	Esophagus			1.68		
Hrt46HR	Heart			0.46		
Kid11Xd	Kidney	28.03	2.34			
Kid109XD	Kidney	6.86	14.02			
Kid10XD	Kidney	27.44	0.03			
Kid124D	Kidney	7.68	2.08			
Kid126XD	Kidney	32.77	3.96			
KID55KD	Kidney			1		
Lvr15XA	Liver	1.75	2.25			
Lv4147L	Liver	8.51	7.91			
Lvr390L	Liver	6.81	1.32			
Liv89LV	Liver			690.81		
Lng354L	Lung	16.02	0			
Lng205L	Lung	2.2	21.32			
LngAC11	Lung	0	14.36			
LngAC39	Lung	17.43	56.6			
Lng315L	Lung	0	7.28			
LngSQ80	Lung	19.12	15.06			
Lng163L	Lung	0	4.21			
LngSQ81	Lung	0	0			
Lng507L	Lung	51.76	45.78			
LNG90LN	Lung			59.43		
OvrG021	Ovary	0.48	0.89			
Ovr206I	Ovary			3.44		
Ovr20GA	Ovary			0		
Ovr18GA	Ovary			7.06		
Ovr3370	Ovary			8.64		
Ovr1230	Ovary			6.51		
OvrC177	Ovary			5.63		
Ovr40G	Ovary			10.85		
Ovr1005O	Ovary	13.69				
Ovr1040O	Ovary	9.66				
Ovr105O	Ovary	29.1				
Ovr130X	Ovary	4.05				

OvrC004	Ovary			2.52		
Ovr63A	Ovary	11.08				
OvrA1B	Ovary	32.27				
Ovr3AOV	Ovary			18.58		
Pan77X	Pancreas	3	3.72			
Pan82XP	Pancreas	21.73	1.45			
Pan92X	Pancreas	20.53	46.03			
Pan35PA	Pancreas					
Pla59PL	Placenta			21.2		
Pro91X	Prostate	2.56	19.24			
Pro109XB	Prostate	0.93	24.15			
Pro134P	Prostate	67.94	56.6			
Pro34B	Prostate	16.65	12.36			
Pro326	Prostate	13.82	34.54			
Pro705P	Prostate				18.12	
Pro784P	Prostate				24.52	
Pro83P	Prostate				33.22	
Pro263C	Prostate				43.33	
Pro10R	Prostate					0
Pro20R	Prostate					19.97
Pro09PR	Prostate			98.67		
Rec21RC	Rectum			0		
Skn248S	Skin	23.15	6.33			
Skn287S	Skin	1.66	32.07			
Skn669S	Skin	22.19	23.83			
Ms184MU	Sktl. Muscle			0.83		
SmInt21XA	Sm. Intestine	47.45	21.79			
SmIntH89A	Sm. Intestine	6.76	16.35			
SmInt20SM	Sm. Intestine	27.43	14.28			
SmInt01SM	Sm. Intestine			7.13		
Spl7GSP	Spleen			4.58		
Sto88S	Stomach	14.72	8.2			
Sto261S	Stomach	10.22	8.61			
Sto288S	Stomach	16.51	0			
StoMT54	Stomach	0.28	2.38			
Sto09ST	Stomach			0.51		
Tst39X	Testis	27.45	8.34			
Tst647T	Testis	36.43	7.8			
Tst663T	Testis	37.5	2.93			
Tst4GTS	Testis			0		
Thy99TM	Thymus			101.8		
Thrd56T	Thyroid	21.99	3.99			
Thrd143N	Thyroid	71.17	23.53			
Thrd270T	Thyroid	0.32	3.43			
Tra16TR	Trachea			25.5		
Utr135XO	Uterus	52.33	49.72			
Utr85XU	Uterus	50.57	65.74			
Utr57UT	Uterus			36.7		

0.00= Negative

The sensitivity for Mam097 expression was calculated for the cancer samples versus normal samples and for the cancer samples versus the expression in the normal adjacent tissue from the same patient. The sensitivity value indicates the percentage of cancer samples that show levels of Mam097 at least 2 fold higher than the normal tissue or

the corresponding normal adjacent form the same patient. Sensitivity data is reported in the table below.

Sensitivity	Breast
≥ 2 fold Up-regulated vs. NAT	53.3%
≥ 2 fold Up-regulated vs. NRM	66.7%

The breast tissue specificity for Mam097 is of 66.7%. This specificity is an indication of the level of breast tissue specific expression of the transcript compared to all the other tissue types tested in our assay. Thus, these experiments indicate Mam097 being useful as a breast cancer diagnostic and/or therapeutic marker.

Altogether, the tissue specificity, plus the mRNA differential expression in the samples tested are believed to make Mam097 a good marker for diagnosing, monitoring, staging, imaging and treating breast cancer.

Primers used for QPCR Expression Analysis of Mam097 are as follows:

SEQ ID NO: 157 (Mam097_probe): CACTTCCTTTAGTTTTGCCCTGG

SEQ ID NO: 158 (Mam097_forward): ATCCTGAATTCTGAGACCATCCA

SEQ ID NO: 159 (Mam097_reverse): GCCTCCAGCACACTCTTCAGT

Mam106 (DEX0432_030.nt.1)

The relative expression level of Mam106 in various tissue samples is included below. Tissue samples include 77 pairs of matching samples, 8 non matched cancer samples, and 36 normal samples, all from various tissues annotated in the table. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Of the normal samples 6 were blood samples which measured the expression level of the BSNA in blood cells. Additionally, 2 prostatitis, and 4 Benign Prostatic Hyperplasia (BPH) samples are included. All the values are compared to bladder cancer sample Bld46XK (calibrator).

The table below contains the relative expression level values for the sample as compared to the calibrator. The table includes the Sample Name, Tissue type, and expression lelvel values for the following samples: Cancer (CAN), Normal Adjacent Tissue (NAT), Normal Tissue (NRM), Benign Prostatic Hyperplasia (BPH), and Prostatitis (PROST).

Sample	Tissue	CAN	NAT	NRM	BPH	PROST
Mam01MA	Breast			0.02		
Mam19DN	Breast	7.35	0.33			

mam355	Breast		0.01			
Mam42DN	Breast					
mam522	Breast	0.08	0.04			
Mam543M	Breast		0.15			
Mam76DN	Breast	0.18	0.11			
Mam781M	Breast	1.89	0.34			
Mam976M	Breast					
MamB011	Breast	0.06	0.21			
mamS516	Breast	0.14	0.02			
MamS570	Breast	3.56	0.88			
mamS621	Breast	1.00	0.01			
MamS699	Breast	1.02	0.25			
MamS854	Breast	1.52	1.49			
MamS997	Breast	0.42	0.36			
Adr48AD	Adrenal			0.16		
Bld23BL	Bladder					
Bld46XK	Bladder	0.94	0.74			
Bld520B	Bladder	2.52	0.31			
BldTR147	Bladder	1.67				
BloB1	Blood			2.32		
BloB11	Blood			5.16		
BloB14	Blood			9.71		
BloB3	Blood			0.44		
BloB5	Blood			1.31		
BloB6	Blood			0.46		
Brn10BR	Brain					
Cvx1ACB	Cervix			8.64		
CvxKS52	Cervix	0.17				
CvxKS83	Cervix	1.42	0.85			
CvxNKS18	Cervix	2.40	0.67			
CvxNKS54	Cervix	16.62	0.77			
CvxNKS81	Cervix	0.62	2.28			
Cln01CL	Colon			0.02		
ClnAS43	Colon	1.65	0.49			
ClnAS98	Colon	4.49	0.45			
ClnCM12	Colon	0.14	0.50			
ClnDC19	Colon	0.70	0.61			
ClnDC63	Colon	1.50	1.23			
ClnRC01	Colon	0.82	0.23			
ClnRS53	Colon	0.23	1.08			
ClnSG27	Colon	0.35	0.40			
ClnTX01	Colon	0.46	0.44			
Endo10479	Endometrium	0.65	3.87			
Endo28XA	Endometrium	2.85	1.68			
Endo3AX	Endometrium	0.33	0.97			
Eso1ES	Esophagus			0.24		
Hrt46HR	Heart					
Kid109XD	Kidney	1.68	0.52			
Kid10XD	Kidney	0.13	0.12			
Kid11Xd	Kidney	0.49	0.28			
Kid124D	Kidney	3.55	0.83			
Kid126XD	Kidney	0.45	0.15			
Kid55KD	Kidney			0.02		
Liv89LV	Liver			0.01		
Lv4147L	Liver	0.23	0.32			
Lvr15XA	Liver	1.59	0.73			
Lvr390L	Liver	0.92	0.59			

Lng163L	Lung	1.62	0.27			
Lng205L	Lung	1.07	0.74			
Lng315L	Lung	1.24	0.78			
Lng354L	Lung	0.59				
Lng507L	Lung	1.22	1.85			
Lng90LN	Lung			0.33		
LngAC11	Lung	0.84	1.19			
LngAC39	Lung	0.71	1.51			
LngSQ80	Lung	6.76	1.16			
LngSQ81	Lung	0.33	0.28			
Ovr1005O	Ovary	0.57				
Ovr1040O	Ovary	0.85				
Ovr105O	Ovary	1.49				
Ovr1230	Ovary			1.07		
Ovr130X	Ovary	1.41				
Ovr18GA	Ovary			0.98		
Ovr206I	Ovary			1.06		
Ovr20GA	Ovary			0.31		
Ovr3370	Ovary			2.17		
Ovr3AOV	Ovary			0.12		
Ovr40G	Ovary			0.99		
Ovr63A	Ovary	0.50				
OvrA1B	Ovary	2.25				
OvrC004	Ovary			3.96		
OvrC177	Ovary			0.47		
OvrG021	Ovary	1.20	1.56			
Pan35PA	Pancreas			0.30		
Pan77X	Pancreas	1.43	0.95			
Pan82XP	Pancreas	0.78	9.90			
Pan92X	Pancreas	1.96	1.31			
Pla59PL	Placenta			0.74		
Pro09PR	Prostate			0.45		
Pro109XB	Prostate	1.85	22.06			
Pro10R	Prostate					
Pro134P	Prostate	7.81	11.88			
Pro20R	Prostate					
Pro263C	Prostate				1.25	
Pro326	Prostate	6.02	2.04			
Pro34B	Prostate	20.81	9.46			
Pro705P	Prostate				2.62	
Pro784P	Prostate				7.35	
Pro83P	Prostate				1.13	
Pro91X	Prostate	3.22	3.76			
Rec21RC	Rectum					
Skn248S	Skin	2.91	0.59			
Skn287S	Skin	1.34	5.85			
Skn669S	Skin	0.21				
Ms184MU	Sktl. Muscle					
SmInt01SM	Sm. Intestine			1.07		
SmInt20SM	Sm. Intestine		0.90			
SmInt21XA	Sm. Intestine	1.61	0.80			
SmIntH89A	Sm. Intestine	0.74	0.45			
Spl7GSP	Spleen			0.09		
Sto09ST	Stomach			0.20		
Sto261S	Stomach	2.80	1.72			
Sto288S	Stomach	0.55	3.35			
Sto88S	Stomach	1.61	2.15			

StoMT54	Stomach	1.04	1.75			
Tst39X	Testis	0.58	2.06			
Tst4GTS	Testis					
Tst647T	Testis	1.74	0.32			
Tst663T	Testis	1.69				
Thy99TM	Thymus					
Thrd143N	Thyroid	2.30	1.04			
Thrd270T	Thyroid	0.69	0.62			
Thrd56T	Thyroid	4.42	1.06			
Tra16TR	Trachea			3.46		
Utr135XO	Uterus	0.24	1.66			
Utr57UT	Uterus			3.03		
Utr85XU	Uterus	2.33	1.95			

0.00= Negative

The sensitivity for Mam106 expression was calculated for the cancer samples versus normal samples and for the cancer samples versus the expression in the normal adjacent tissue from the same patient. The sensitivity value indicates the percentage of cancer samples that show levels of Mam106 at least 2 fold higher than the normal tissue or the corresponding normal adjacent form the same patient. Sensitivity data is reported in the table below.

Sensitivity	Breast
≥ 2 fold Up-regulated vs. NAT	50%
≥ 2 fold Up-regulated vs. NRM	79%

The breast tissue specificity for Mam106 is of 10%. This specificity is an indication of the level of breast tissue specific expression of the transcript compared to all the other tissue types tested in our assay. Thus, these experiments indicate Mam106 being useful as a breast cancer diagnostic and/or therapeutic marker.

Altogether, the tissue specificity, plus the mRNA differential expression in the samples tested are believed to make Mam106 a good marker for diagnosing, monitoring, staging, imaging and treating breast cancer.

Primers used for QPCR Expression Analysis of Mam106 are as follows:

SEQ ID NO: 160 (Mam106_probe): AGCCGGAGGAGATGTGGCTCTACCG

SEQ ID NO: 161 (Mam106_forward): CCGCTTCCCAGAGACTCATC

SEQ ID NO: 162 (Mam106_reverse): GCACAAACATCGGCTTGGT

<u>Mam096 (DEX0432_035.nt.1; DEX0432_036.nt.1 - DEX0432_036.nt.48)</u>

The relative expression level of Mam096 in various tissue samples is included below. Tissue samples include 77 pairs of matching samples, 8 non matched cancer samples, and 36 normal samples, all from various tissues annotated in the table. A

matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Of the normal samples 6 were blood samples which measured the expression level of the BSNA in blood cells. Additionally, 2 prostatitis, and 4 Benign Prostatic Hyperplasia (BPH) samples are included. All the values are compared to breast cancer sample mamS621 (calibrator).

The table below contains the relative expression level values for the sample as compared to the calibrator. The table includes the Sample Name, Tissue type, and expression lelvel values for the following samples: Cancer (CAN), Normal Adjacent Tissue (NAT), Normal Tissue (NRM), Benign Prostatic Hyperplasia (BPH), and Prostatitis (PROST).

Sample	Tissue	CAN	NAT	NRM	BPH	PROST
Mam01MA	Breast			1.75		
Mam19DN	Breast	3.84	0.00			
mam355	Breast	1.00	0.004			
Mam42DN	Breast		0.00			
mam522	Breast	0.04	0.00			
Mam543M	Breast	2.26	0.00			
Mam76DN	Breast	0.06	0.00			
Mam781M	Breast		1.53			
Mam976M	Breast		0.00			
MamB011	Breast	0.50	0.62			
mamS516	Breast		0.00			
MamS570	Breast	13.56	0.00			
mamS621	Breast	1.00	0.00			
MamS699	Breast	0.93	0.00			
MamS854	Breast		0.00			
MamS997	Breast		0.00			
Adr48AD	Adrenal			0.00		
Bld23BL	Bladder					
Bld46XK	Bladder	0.00	52.72			
Bld520B	Bladder	0.00	0.00			
BldTR147	Bladder	9.39	0.00			
BloB1	Blood			0.00		
BloB11	Blood			0.00		
BloB14	Blood			0.00		
BloB3	Blood			0.00		
BloB5	Blood			0.00		
BloB6	Blood			0.00		
Brn10BR	Brain			0.00		
Cvx1ACB	Cervix			0.00		
CvxKS52	Cervix	0.00	0.00			
CvxKS83	Cervix	0.00	0.00			
CvxNKS18	Cervix	4.38	0.00			
CvxNKS54	Cervix	7.72	0.00			
CvxNKS81	Cervix	0.00	0.00			
Cln01CL	Colon			0.20		
ClnAS43	Colon	6.96	1.64			
ClnAS98	Colon	3.04	0.00			

ClnCM12	Colon	0.55	0.00			
ClnDC19	Colon	0.65	0.00			
ClnDC63	Colon	2.07	0.00			
ClnRC01	Colon	0.00	0.00			
ClnRS53	Colon	0.00	0.00			
ClnSG27	Colon	2.55	3.07			
ClnTX01	Colon	0.00	2.41			
Endo10479	Endometrium	5.22	0.00			
Endo28XA	Endometrium	7.02	9.74			
Endo3AX	Endometrium	0.00	0.00			
Eso1ES	Esophagus			0.00		
Hrt46HR	Heart			0.00		
Kid109XD	Kidney	1.25	1.31			
Kid10XD	Kidney	0.52	0.00			
Kid11Xd	Kidney	0.22	3.08			
Kid124D	Kidney	0.00	4.62			
Kid126XD	Kidney	0.00	3.64			
Kid55KD	Kidney			0.04		
Liv89LV	Liver			0.00		
Lv4147L	Liver	0.00	0.00			
Lvr15XA	Liver	0.00	0.00			
Lvr390L	Liver	18.15	0.00			
Lng163L	Lung	0.00	0.00			
Lng205L	Lung	3.50	91.36			
Lng315L	Lung	0.00	0.00			
Lng354L	Lung	0.00	0.00			
Lng507L	Lung	67.15	10.06			
Lng90LN	Lung			0.21		
LngAC11	Lung	21.28	19.12			
LngAC39	Lung	61.63	0.00			
LngSQ80	Lung	5.45	21.39			
LngSQ81	Lung	15.42	131.51			
Ovr1005O	Ovary	73.79				
Ovr1040O	Ovary	6.69				
Ovr105O	Ovary	54.02				
Ovr1230	Ovary			0.00		
Ovr130X	Ovary	118.65				
Ovr18GA	Ovary			0.00		
Ovr206I	Ovary			1.22		
Ovr20GA	Ovary			0.00		
Ovr3370	Ovary			0.00		
Ovr3AOV	Ovary			0.00		
Ovr40G	Ovary			0.00		
Ovr63A	Ovary	0.00				
OvrA1B	Ovary	1257.69				
OvrC004	Ovary			0.00		
OvrC177	Ovary			0.00		
OvrG021	Ovary	28.86	0.00			
Pan35PA	Pancreas			0.47		
Pan77X	Pancreas	0.00	0.00			
Pan82XP	Pancreas	0.18	0.00			
Pan92X	Pancreas	292.34	0.00			
Pla59PL	Placenta			0.00		
Pro09PR	Prostate			0.04		
Pro109XB	Prostate	0.00	0.00			
Pro10R	Prostate					0.00
Pro134P	Prostate	0.00	0.00			

Pro20R	Prostate					0.00
Pro263C	Prostate				0.00	
Pro326	Prostate	0.00	0.00			
Pro34B	Prostate	0.00	0.00			
Pro705P	Prostate				0.00	
Pro784P	Prostate				0.00	
Pro83P	Prostate				0.00	
Pro91X	Prostate	0.00	0.00			
Rec21RC	Rectum			0.00		
Skn248S	Skin	0.00	0.00			
Skn287S	Skin	0.00	0.00			
Skn669S	Skin	0.00	0.00			
Ms184MU	Sktl. Muscle			0.00		
SmInt01SM	Sm. Intestine			0.00		
SmInt20SM	Sm. Intestine	0.00	0.00			
SmInt21XA	Sm. Intestine	0.00	0.00			
SmIntH89A	Sm. Intestine	12.43	0.00			
Spl7GSP	Spleen			0.27		
Sto09ST	Stomach			0.20		
Sto261S	Stomach	0.00	0.00			
Sto288S	Stomach	32.51	0.00			
Sto88S	Stomach	0.00	5.74			
StoMT54	Stomach	7.37	0.00			
Tst39X	Testis	0.00	0.00			
Tst4GTS	Testis			0.00		
Tst647T	Testis	0.00	0.00			
Tst663T	Testis	0.00	0.00			
Thy99TM	Thymus			0.00		
Thrd143N	Thyroid	0.00	0.00			
Thrd270T	Thyroid	1.32	1.77			
Thrd56T	Thyroid	0.00	0.00			
Tra16TR	Trachea			0.00		
Utr135XO	Uterus	1.92	2.26			
Utr57UT	Uterus			0.00		
Utr85XU	Uterus	10.00	0.00			

0.00= Negative

The sensitivity for Mam096 expression was calculated for the cancer samples versus normal samples and for the cancer samples versus the expression in the normal adjacent tissue from the same patient. The sensitivity value indicates the percentage of cancer samples that show levels of Mam096 at least 2 fold higher than the normal tissue or the corresponding normal adjacent form the same patient. Sensitivity data is reported in the table below.

Sensitivity	Breast
≥ 2 fold Up-regulated vs. NAT	53%
≥ 2 fold Up-regulated vs. NRM	13%

The breast tissue specificity for Mam096 is of 49%. This specificity is an indication of the level of breast tissue specific expression of the transcript compared to all

the other tissue types tested in our assay. Thus, these experiments indicate Mam096 being useful as a breast cancer diagnostic and/or therapeutic marker.

Altogether, the tissue specificity, plus the mRNA differential expression in the samples tested are believed to make Mam096 a good marker for diagnosing, monitoring, staging, imaging and treating breast cancer.

Primers used for QPCR Expression Analysis of Mam096 are as follows:

SEQ ID NO: 163 (Mam096_probe): AGAGAGACATTTCTGAAATGGCTGTCT

SEQ ID NO: 164 (Mam096_forward): CCCAGCACCGACTACTACCAA

SEQ ID NO: 165 (Mam096_reverse): AGCTGCCCGTAGTTCTTTCG

Mam103 (DEX0432_038.nt.1)

The relative expression level of Mam103 in various tissue samples is included below. Tissue samples include 74 pairs of matching samples, 7 non matched cancer samples, and 38 normal samples, all from various tissues annotated in the table. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Of the normal samples 6 were blood samples which measured the expression level of the BSNA in blood cells. Additionally, 2 prostatitis, and 4 Benign Prostatic Hyperplasia (BPH) samples are included. All the values are compared to breast normal adjacent sample Mam522 (calibrator).

The table below contains the relative expression level values for the sample as compared to the calibrator. The table includes the Sample Name, Tissue type, and expression lelvel values for the following samples: Cancer (CAN), Normal Adjacent Tissue (NAT), Normal Tissue (NRM), Benign Prostatic Hyperplasia (BPH), and Prostatitis (PROST).

Sample	Tissue	CAN	NAT	NRM	BPH	PROST
Mam01MA	Breast			0.00		
Mam19DN	Breast	0.04	0.00			
Mam42DN	Breast	0.00	0.00			
Mam522	Breast		1.00			
Mam543M	Breast	N/A	0.01			
Mam781M	Breast	0.15	0.02			
Mam976M	Breast	0.05	0.00			
MamS516	Breast	0.20				
MamS570	Breast	0.00	0.00			
MamS699	Breast	0.01	0.00			

MamS854	Breast	0.00	0.00			
MamS997	Breast	0.01	0.00			
Adr48AD	Adrenal			0.00		
Bld23BL	Bladder			0.00		
Bld46XK	Bladder	0.11	0.05			
Bld520B	Bladder	0.11	0.03			
BldTR147	Bladder	0.15	0.00			
BloB1	Blood			1.73		
BloB11	Blood			2.49		
BloB14	Blood			0.42		
BloB3	Blood			0.95		
BloB5	Blood			0.43		
BloB6	Blood			0.42		
Brn10BR	Brain			0.00		
Cvx1ACB	Cervix			0.18		
CvxKS52	Cervix	0.01	0.00			
CvxKS83	Cervix	0.39	0.00			
CvxNKS18	Cervix	0.08	0.08			
CvxNKS54	Cervix	0.09	0.00			
CvxNKS81	Cervix	0.09	0.00			
Cln01CL	Colon			0.00		
ClnAS43	Colon	0.02	0.00			
ClnAS98	Colon	0.02	0.00			
ClnCM12	Colon	0.00	0.01			
ClnDC19	Colon	0.01	0.09			
ClnDC63	Colon	0.01	0.02			
ClnRC01	Colon	0.00	0.00			
ClnRS53	Colon	0.00	0.01			
ClnSG27	Colon	0.00	0.02			
ClnTX01	Colon	0.01	0.00			
Endo10479	Endmetrium	0.15	0.00			
Endo28XA	Endmetrium	0.03	0.03			
Endo3AX	Endmetrium	0.02	0.05			
Eso1ES	Esophagus			0.00		
Hrt46HR	Heart			0.00		
Kid109XD	Kidney	0.02	0.02			
Kid10XD	Kidney	0.00	0.00			
Kid11Xd	Kidney	0.11	0.04			
Kid124D	Kidney	0.02	0.00			
Kid126XD	Kidney	0.05	0.00			
Kid55KD	Kidney			0.02		
Liv89LV	Liver			0.00		
Lv4147L	Liver	0.00	0.00			
Lvr15XA	Liver	0.00	0.01			
Lvr390L	Liver	0.09	0.00			
Lng163L	Lung	0.05	0.00			
Lng205L	Lung	0.00	0.00			

Lng315L	Lung	0.02	0.00			
Lng354L	Lung	0.00	0.00			
Lng507L	Lung	0.03	0.00			
Lng90LN	Lung			0.14		
LngAC11	Lung	0.02	0.00			
LngAC39	Lung	0.02	0.00			
LngSQ80	Lung	0.00	0.04			
LngSQ81	Lung	0.01	0.01			
Ovr1005O	Ovary	0.02				
Ovr1040O	Ovary	0.00				
Ovr105O	Ovary	0.02				
Ovr1230	Ovary			0.00		
Ovr130X	Ovary	0.03				
Ovr18GA	Ovary			0.12		
Ovr206I	Ovary			0.04		
Ovr20GA	Ovary			0.03		
Ovr3370	Ovary			0.07		
Ovr3AOV	Ovary			0.00		
Ovr40G	Ovary			0.04		
Ovr63A	Ovary	0.02				
OvrA1B	Ovary	0.08				
OvrC004	Ovary			0.22		
OvrC177	Ovary			0.06		
OvrG021	Ovary	0.02	0.00			
Pan35PA	Pancreas			0.00		
Pan77X	Pancreas	0.00	0.00			
Pan82XP	Pancreas	0.02	0.81			
Pan92X	Pancreas	0.00	0.02			
Pla59PL	Placenta			0.02		
Pro09PR	Prostate			0.00		
Pro109XB	Prostate	0.01	0.04			
Pro10R	Prostate					0.00
Pro134P	Prostate	0.04	0.05			
Pro20R	Prostate					N/A
Pro263C	Prostate				0.06	
Pro326	Prostate	0.01	0.01			
Pro34B	Prostate	0.00	0.00			
Pro705P	Prostate				0.00	
Pro784P	Prostate				0.03	
Pro83P	Prostate				0.00	
Pro91X	Prostate	0.00	0.00			
Rec21RC	Rectum			0.00		
Skn248S	Skin	0.14	0.00			
Skn287S	Skin	0.01	0.00			
Skn669S	Skin	0.00	0.00			
Ms184MU	Sktl. Muscle			0.00		
SmInt01SM	Sm. Instestine			0.00		

SmInt20SM	Sm. Instestine	0.00	0.00			
SmInt21XA	Sm. Instestine	0.00	0.00			
SmIntH89A	Sm. Instestine	0.05	0.02			
Spl7GSP	Spleen			0.02		
Sto09ST	Stomach			0.00		
Sto261S	Stomach	0.08	0.00			
Sto288S	Stomach	0.02	1.12			
Sto88S	Stomach	0.00	0.00			
StoMT54	Stomach	0.02	0.02			
Tst39X	Testis	0.16	0.01			
Tst4GTS	Testis			0.00		
Tst647T	Testis	0.02	0.01			
Tst663T	Testis	0.06	0.01			
Thy99TM	Thymus			0.00		
Thrd143N	Thyroid	0.01	0.00			
Thrd270T	Thyroid	0.01	0.00			
Thrd56T	Thyroid	0.00	0.22			
Tra16TR	Trachea			0.00		
Utr135XO	Uterus	0.01	0.01			
Utr57UT	Uterus			0.00		
Utr85XU	Uterus	0.03	0.02			

0.00= Negative

The sensitivity for Mam103 expression was calculated for the cancer samples versus normal samples and for the cancer samples versus the expression in the normal adjacent tissue from the same patient. The sensitivity value indicates the percentage of cancer samples that show levels of Mam103 at least 2 fold higher than the normal tissue or the corresponding normal adjacent form the same patient. Sensitivity data is reported in the table below.

Sensitivity	Breast
≥ 2 fold Up-regulated vs. NAT	79%
≥ 2 fold Up-regulated vs. NRM	63%

The breast tissue specificity for Mam103 is of 42%. This specificity is an indication of the level of breast tissue specific expression of the transcript compared to all the other tissue types tested in our assay. Thus, these experiments indicate Mam103 being useful as a breast cancer diagnostic and/or therapeutic marker.

Altogether, the tissue specificity, plus the mRNA differential expression in the samples tested are believed to make Mam103 a good marker for diagnosing, monitoring, staging, imaging and treating breast cancer.

Primers used for QPCR Expression Analysis of Mam103 are as follows:

SEQ ID NO: 166 (Mam103_probe): CTGAAAGCAGGTCACCCCTGAGATCCT

SEQ ID NO: 167 (Mam103_forward): CAGAGCTTGGCCAGGTTCTAA

SEQ ID NO: 169 (Mam103_reverse): TGCTAGGGTGCCCCTCTGT

Mam098 (DEX0432_044.nt.1)

The relative expression level of Mam098 in various tissue samples is included below. Tissue samples include 6 pairs of matching samples, and 11 normal samples, all from various tissues annotated in the table. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. All the values are compared to breast cancer sample Mam76DN (calibrator).

The table below contains the relative expression level values for the sample as compared to the calibrator. The table includes the Sample Name, Tissue type, and expression lelvel values for the following samples: Cancer (CAN), Normal Adjacent Tissue (NAT), Normal Tissue (NRM).

Sample	Tissue	CAN	NAT	NRM
Mam01MA	Mammary			0.03
mam355	Mammary	0.07	0.01	
MamB011	Mammary	0.00	0.22	
mamS621	Mammary	0.47	0.00	
mamS516	Mammary	0.19	0.00	
mam522	Mammary	0.45	0.01	
Mam76DN	Mammary	1.00	0.05	
Bld23BL	Bladder			0.00
Cln01CL	Colon			0.01
Kid55KD	Kidney			0.00
Liv89LV	Liver			0.01
Lng90LN	Lung			0.01
Ovr3AOV	Ovary			0.00
Pan35PA	Pancreas			0.00
Pro09PR	Prostate			0.06
Spl7GSP	Spleen			0.03
Sto09ST	Stomach			0.00

0.00= Negative

The tissue specificity, plus the mRNA differential expression in the samples tested are believed to make Mam098 a good marker for diagnosing, monitoring, staging, imaging and treating breast cancer.

Primers used for QPCR Expression Analysis of Mam098 are as follows:

SEQ ID NO: 169 (Mam098_probe): CCTTTAGGGCCTGGGACAACCACG

SEQ ID NO: 170 (Mam098_forward): TGGATAACAAGCCCACAAATGA

SEQ ID NO: 171 (Mam098_reverse): CCTCTAGTTCCAGCCCCTTTTAG

Example 3: Protein Expression

The BSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the BSNA is subcloned in pET-21d for expression in *E. coli.* In addition to the BSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of BSNA, and six histidines, flanking the COOH-terminus of the coding sequence of BSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of BSP is achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that are separated from total cell lysate were incubated with a nickle chelating resin. The column is packed and washed with five column volumes of wash buffer. BSP is eluted stepwise with various concentration imidazole buffers.

Example 4: Fusion Proteins

The human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if

the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/1 of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al., Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

The polypeptides of the present invention were analyzed and the following attributes were identified; specifically, epitopes, post translational modifications, signal peptides and transmembrane domains. Antigenicity (Epitope) prediction was performed through the antigenic module in the EMBOSS package. Rice, P., EMBOSS: The European Molecular Biology Open Software Suite, *Trends in Genetics* 16(6): 276-277

(2000). The antigenic module predicts potentially antigenic regions of a protein sequence, using the method of Kolaskar and Tongaonkar. Kolaskar, AS and Tongaonkar, PC., A semi-empirical method for prediction of antigenic determinants on protein antigens, *FEBS Letters* 276: 172-174 (1990). Examples of post-translational modifications (PTMs) and other motifs of the ***XSP***s of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. The PTMs and other motifs were predicted by using the ProSite Dictionary of Proteins Sites and Patterns (Bairoch *et al.*, *Nucleic Acids Res.* 25(1):217-221 (1997)), the following motifs, including PTMs, were predicted for the ***XSP***s of the invention. The signal peptides were detected by using the SignalP 2.0, *see* Nielsen *et al.*, *Protein Engineering* 12, 3-9 (1999). Prediction of transmembrane helices in proteins was performed by the application TMHMM 2.0, "currently the best performing transmembrane prediction program", according to authors (Krogh *et al.*, *Journal of Molecular Biology*, 305(3):567-580, (2001); Moller *et al.*, *Bioinformatics*, 17(7):646-653, (2001); Sonnhammer, *et al.*, *A hidden Markov model for predicting transmembrane helices in protein sequences* in Glasgow, *et al.* Ed. Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press. The PSORT II program may also be used to predict cellular localizations. Horton et al., *Intelligent Systems for Molecular Biology* 5: 147-152 (1997). The table below includes the following sequence annotations: Signal peptide presence; TM (number of membrane domain, topology in orientation and position); Amino acid location and antigenic index (location, AI score, length); PTM and other motifs (type, amino acid residue locations); and functional domains.

AA SEQ ID	Sig P	TMHMM	Antigenicity	PTM	Domain
DEX0432_005.aa.1			160-170, 1.17, 11; 96-113, 1.09, 18	Amidation 3-6; Camp_Phospho_Site 133-136; Glycosaminoglycan 71-74; Myristyl 60-65;61-66;70-75;72-77;113-118;115-120; Pkc_Phospho_Site 3-5;47-49;132-134;140-142;	
DEX0432_010.aa.1				Asn_Glycosylation 3-6; Myristyl 8-13;	
DEX0432_013.aa			404-414, 1.14, 11;	Amidation 352-355;	

.1			312-324, 1.02, 13; 466-476, 1.00, 11	Asn_Glycosylation 609-612; Camp_Phospho_Site 216-219;355-358; Ck2_Phospho_Site 80-83;156-159;298-301;330-333;399-402;403-406;416-419;429-432;537-540;551-554;596-599; Glycosaminoglycan 417-420; Myristyl 43-48;85-90;370-375;374-379;412-417;424-429;492-497; Pkc_Phospho_Site 9-11;58-60;143-145;268-270;298-300;409-411;471-473;481-483;482-484;510-512; Protein_Kinase_Atp 68-91; Protein_Kinase_St 181-193;	
DEX0432_015.aa.1				Asn_Glycosylation 13-16; Ck2_Phospho_Site 15-18; Myristyl 36-41;85-90; Pkc_Phospho_Site 73-75;89-91; Tyr_Phospho_Site 80-87;81-87;	
DEX0432_016.aa.1		1 i21-43o		Amidation 76-79; Asn_Glycosylation 7-10;136-139; Ck2_Phospho_Site 138-141; Myristyl 111-116;159-164;206-211; Pkc_Phospho_Site 47-49;61-63;97-99;194-196;210-212; Ribosomal_L6_2 191-212; Tyr_Phospho_Site 201-208;202-208;	
DEX0432_017.aa.1				Asn_Glycosylation 22-25;	
DEX0432_018.aa.1			27-49, 1.23, 23	Pkc_Phospho_Site 3-5;6-8;39-41; Ribosomal_L39e 30-46;	
DEX0432_019.aa		1 i28-			

.1		50o			
DEX0432 _021.aa .1		1 i96-118o		Myristyl 57-62; Pkc_Phospho_Site 4-6;	
DEX0432 _023.aa .1			79-94, 1.06, 16	Amidation 10-13;79-82; Asn_Glycosylation 104-107; Myristyl 19-24;66-71;76-81;87-92; Pkc_Phospho_Site 34-36; Tyr_Phospho_Site 93-100;	
DEX0432 _025.aa .1			23-34, 1.04, 12	Camp_Phospho_Site 64-67; Ck2_Phospho_Site 5-8; Myristyl 54-59;72-77;76-81; Pkc_Phospho_Site 5-7;30-32;50-52;55-57;	
DEX0432 _026.aa .1				Ck2_Phospho_Site 65-68;90-93;129-132; Myristyl 72-77;106-111;126-131; Pkc_Phospho_Site 19-21;89-91;	
DEX0432 _031.aa .1		3 i13-35o45 - 67i13 3-155o	155-168, 1.24, 14; 79-95, 1.09, 17	Asn_Glycosylation 81-84;105-108; Ck2_Phospho_Site 145-148; Myristyl 4-9;15-20;31-36;141-146;151-156; Pkc_Phospho_Site 10-12;85-87;155-157; Prokar_Lipoprotei n 52-62;54-64;	
DEX0432 _033.aa .1			15-29, 1.14, 15	Asn_Glycosylation 11-14; Ck2_Phospho_Site 23-26;51-54; Myristyl 48-53;	
DEX0432 _035.aa .1			56-116, 1.03, 61	Ck2_Phospho_Site 43-46;55-58;92-95;97-100; Myristyl 62-67;105-110;108-113; Pkc_Phospho_Site 40-42;43-45;94-96;	
DEX0432 _036.aa .2	y	1 - o1160 - 1182i	1154-1189,1.349; 1097-1107,1.225;	Atpase_Alpha_Beta 959-968; Asn_Glycosylation 963-966, 981-984;	PS00152, SEA, SEA, PRO_RICH, SEA,

			6-24,1.22; 1132-1152,1.158; 98-111,1.149; 889-901,1.134; 1034-1053,1.133; 949-962,1.12; 56-65,1.12; 1017-1029,1.112; 970-977,1.105; 909-922,1.096; 77-85,1.092; 369-381,1.088; 569-581,1.088; 509-521,1.088; 429-441,1.088; 449-461,1.088; 669-681,1.088; 389-401,1.088; 609-621,1.088; 349-361,1.088; 469-481,1.088; 249-261,1.088; 129-141,1.088; 709-721,1.088; 329-341,1.088; 209-221,1.088; 489-501,1.088; 269-281,1.088; 409-421,1.088; 829-841,1.088; 649-661,1.088; 189-201,1.088; 769-	Ck2_Phospho_Site 52-55, 125-128, 145-148, 165-168, 185-188, 205-208, 225-228, 245-248, 265-268, 285-288, 305-308, 325-328, 345-348, 365-368, 385-388, 405-408, 425-428, 445-448, 465-468, 485-488, 505-508, 525-528, 545-548, 565-568, 585-588, 605-608, 625-628, 645-648, 665-668, 685-688, 705-708, 725-728, 745-748, 765-768, 785-788, 805-808, 825-828, 845-848, 865-868, 885-888, 905-908, 926-929, 946-949; Glycosaminoglycan 911-914; Myristyl 28-33, 72-77, 74-79, 80-85, 94-99, 100-105, 969-974, 973-978; Pkc_Phospho_Site 45-47, 54-56, 984-986;	

			781,1.088; 849-861,1.088; 589-601,1.088; 869-881,1.088; 930-942,1.088; 529-541,1.088; 789-801,1.088; 729-741,1.088; 309-321,1.088; 289-301,1.088; 809-821,1.088; 549-561,1.088; 229-241,1.088; 149-161,1.088; 629-641,1.088; 749-761,1.088; 689-701,1.088; 169-181,1.088; 1076-1085,1.088; 1216-1223,1.086; 1113-1119,1.082; 1241-1253,1.08; 989-1000,1.077; 115-121,1.065; 1087-1093,1.058; 42-48,1.058; 1067-1073,1.047		
DEX0432_036.aa.3	y	1 - o551-573i	545-580,1.349; 426-450,1.225; 522-543,1.224; 6-24,1.22; 583-607,1.195;	Atpase_Alpha_Beta 219-228; Amidation 718-721; Asn_Glycosylation 223-226, 241-244, 295-298, 321-324, 482-485;	PS00152, PRO_RICH, SEA,

			352-409,1.172; 488-513,1.156; 98-111,1.149; 294-313,1.133; 209-222,1.12; 56-65,1.12; 277-289,1.112; 336-350,1.111; 230-237,1.105; 648-656,1.102; 666-676,1.102; 77-85,1.092; 170-182,1.088; 190-202,1.088; 149-162,1.088; 129-141,1.088; 634-641,1.086; 475-484,1.083; 456-462,1.082; 249-260,1.077; 115-121,1.065; 42-48,1.058; 327-333,1.047	Ck2_Phospho_Site 52-55, 125-128, 145-148, 166-169, 186-189, 206-209, 276-279, 322-325, 323-326, 351-354, 484-487, 646-649, 651-654; Glycosaminoglycan 520-523, 549-552; Myristyl 28-33, 72-77, 74-79, 80-85, 94-99, 100-105, 229-234, 233-238, 383-388, 394-399, 416-421, 519-524, 523-528, 593-598, 711-716, 715-720, 726-731; Pkc_Phospho_Site 45-47, 54-56, 244-246, 275-277, 417-419, 520-522, 526-528, 648-650, 730-732;	
DEX0432_036.aa.4	y	1 - o428-450i	422-457,1.349; 365-375,1.225; 6-21,1.22; 400-420,1.158; 107-120,1.149; 302-321,1.133; 217-230,1.12; 65-74,1.12; 285-297,1.112; 26-33,1.111; 238-245,1.105; 86-94,1.092; 198-210,1.088;	Atpase_Alpha_Beta 227-236; Asn_Glycosylation 231-234, 249-252, 303-306, 329-332, 407-410; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 194-197, 214-217, 284-287, 330-333, 331-334, 409-412, 414-417, 496-499, 501-504; Glycosaminoglycan 426-429; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 237-242, 241-246,	PS00152, SEA, SEA, PRO_RICH, SEA,

			178-190,1.088; 138-150,1.088; 158-170,1.088; 344-353,1.088; 484-491,1.086; 381-387,1.082; 509-521,1.08; 257-268,1.077; 124-130,1.065; 355-361,1.058; 51-57,1.058; 335-341,1.047	358-363, 509-514, 512-517; Pkc_Phospho_Site 54-56, 63-65, 252-254, 283-285, 498-500;	
DEX0432_036.aa.5	y	1 - o419-441i	413-448,1.349; 356-366,1.225; 6-24,1.22; 391-411,1.158; 98-111,1.149; 293-312,1.133; 208-221,1.12; 56-65,1.12; 276-288,1.112; 229-236,1.105; 77-85,1.092; 189-201,1.088; 129-141,1.088; 169-181,1.088; 149-161,1.088; 335-344,1.088; 475-482,1.086; 372-378,1.082; 500-512,1.08; 248-259,1.077; 115-121,1.065; 346-352,1.058; 42-48,1.058; 326-332,1.047	Atpase_Alpha_Beta 218-227; Asn_Glycosylation 222-225, 240-243, 294-297, 320-323, 398-401; Ck2_Phospho_Site 52-55, 125-128, 145-148, 165-168, 185-188, 205-208, 275-278, 321-324, 322-325, 400-403, 405-408, 487-490, 492-495; Glycosaminoglycan 417-420; Myristyl 28-33, 72-77, 74-79, 80-85, 94-99, 100-105, 228-233, 232-237, 349-354, 500-505, 503-508; Pkc_Phospho_Site 45-47, 54-56, 243-245, 274-276, 489-491;	PS00152, SEA, SEA, PRO_RICH, SEA,

DEX0432 _036.aa .6	N	0 -o	72-91,1.23; 93-106,1.204; 53-67,1.116; 14-25,1.088; 33-45,1.088; 5-12,1.055	Amidation 74-77; Ck2_Phospho_Site 29-32, 49-52, 94-97; Pkc_Phospho_Site 74-76, 98-100; Tyr_Phospho_Site 103-109;	
DEX0432 _036.aa .7	Y	0 -o	6-21,1.22; 151-171,1.182; 107-120,1.149; 65-74,1.12; 26-33,1.111; 86-94,1.092; 143-149,1.071; 124-130,1.065; 51-57,1.058	Ck2_Phospho_Site 61-64, 134-137; Glycosaminoglycan 153-156; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 152-157, 154-159, 169-174; Pkc_Phospho_Site 54-56, 63-65; Prokar_Lipoprotein 149-159;	
DEX0432 _036.aa .8	y	1 - o379-401i	373-408,1.349; 316-326,1.225; 6-24,1.22; 351-371,1.158; 98-111,1.149; 253-272,1.133; 168-181,1.12; 56-65,1.12; 236-248,1.112; 189-196,1.105; 77-85,1.092; 149-161,1.088; 129-141,1.088; 295-304,1.088; 435-442,1.086; 332-338,1.082; 460-472,1.08; 208-219,1.077; 115-121,1.065; 306-312,1.058; 42-48,1.058; 286-292,1.047	Atpase_Alpha_Beta 178-187; Asn_Glycosylation 182-185, 200-203, 254-257, 280-283, 358-361; Ck2_Phospho_Site 52-55, 125-128, 145-148, 165-168, 235-238, 281-284, 282-285, 360-363, 365-368, 447-450, 452-455; Glycosaminoglycan 377-380; Myristyl 28-33, 72-77, 74-79, 80-85, 94-99, 100-105, 188-193, 192-197, 309-314, 460-465, 463-468; Pkc_Phospho_Site 45-47, 54-56, 203-205, 234-236, 449-451;	PS00152, SEA, SEA, SEA,
DEX0432 _036.aa .10	Y	0 -o	163-174,1.22; 4-13,1.198; 208-218,1.191;	Camp_Phospho_Site 189-192; Ck2_Phospho_Site 47-50, 107-110,	C2, C2, C2_DOMAIN_2, PR00399,

			221-228,1.177; 43-73,1.167; 116-153,1.151; 15-28,1.145; 31-41,1.138; 78-90,1.109; 92-101,1.101; 198-205,1.07; 188-194,1.046	114-117, 116-119, 133-136, 161-164, 218-221; Myristyl 45-50, 66-71; Pkc_Phospho_Site 42-44, 143-145, 192-194; Tyr_Phospho_Site 96-103, 194-201;	
DEX0432_036.aa.11	N	0 -o	25-55,1.167; 60-96,1.161; 13-23,1.138; 4-10,1.107; 98-104,1.056	Ck2_Phospho_Site 29-32; Myristyl 27-32, 48-53, 101-106; Pkc_Phospho_Site 24-26;	
DEX0432_036.aa.12	y	1 - o388-410i	382-417,1.349; 325-335,1.225; 6-21,1.22; 360-380,1.158; 107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 158-170,1.088; 138-150,1.088; 304-313,1.088; 444-451,1.086; 341-347,1.082; 469-481,1.08; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292, 367-370; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372, 374-377, 456-459, 461-464; Glycosaminoglycan 386-389; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323, 469-474, 472-477; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245, 458-460;	PS00152, SEA, SEA, SEA,
DEX0432_036.aa.13	y	1 - o367-389i	361-396,1.349; 304-314,1.225; 6-20,1.22; 339-359,1.158;	Atpase_Alpha_Beta 166-175; Asn_Glycosylation 170-173, 188-191, 242-245, 268-271, 346-349;	PS00152, SEA, SEA, SEA,

			86-99,1.149; 241-260,1.133; 156-169,1.12; 44-53,1.12; 224-236,1.112; 177-184,1.105; 65-73,1.092; 137-149,1.088; 117-129,1.088; 283-292,1.088; 423-430,1.086; 320-326,1.082; 448-460,1.08; 196-207,1.077; 103-109,1.065; 294-300,1.058; 30-36,1.058; 274-280,1.047	Ck2_Phospho_Site 24-27, 40-43, 113-116, 133-136, 153-156, 223-226, 269-272, 270-273, 348-351, 353-356, 435-438, 440-443; Glycosaminoglycan 365-368; Myristyl 60-65, 62-67, 68-73, 82-87, 88-93, 176-181, 180-185, 297-302, 448-453, 451-456; Pkc_Phospho_Site 33-35, 42-44, 191-193, 222-224, 437-439;	
DEX0432_036.aa.14	N	1 - o428-450i	422-457,1.349; 19-52,1.265; 365-375,1.225; 400-420,1.158; 147-160,1.149; 302-321,1.133; 217-230,1.12; 105-114,1.12; 63-73,1.114; 285-297,1.112; 238-245,1.105; 126-134,1.092; 198-210,1.088; 178-190,1.088; 344-353,1.088; 484-491,1.086; 381-387,1.082; 509-521,1.08;	Atpase_Alpha_Beta 227-236; Asn_Glycosylation 231-234, 249-252, 303-306, 329-332, 407-410; Ck2_Phospho_Site 101-104, 174-177, 194-197, 214-217, 284-287, 330-333, 331-334, 409-412, 414-417, 496-499, 501-504; Glycosaminoglycan 426-429; Myristyl 22-27, 77-82, 121-126, 123-128, 129-134, 143-148, 149-154, 237-242, 241-246, 358-363, 509-514, 512-517; Pkc_Phospho_Site 94-96, 103-105, 252-254, 283-285, 498-500;	PS00152, SEA, SEA, SEA,

			257-268,1.077; 164-170,1.065; 355-361,1.058; 91-97,1.058; 335-341,1.047		
DEX0432_036.aa.15	y	1 - o339-361i	333-368,1.349; 276-286,1.225; 6-24,1.22; 311-331,1.158; 98-111,1.149; 213-232,1.133; 128-141,1.12; 56-65,1.12; 196-208,1.112; 149-156,1.105; 77-85,1.092; 255-264,1.088; 395-402,1.086; 292-298,1.082; 420-432,1.08; 168-179,1.077; 115-121,1.065; 266-272,1.058; 42-48,1.058; 246-252,1.047	Atpase_Alpha_Beta 138-147; Asn_Glycosylation 142-145, 160-163, 214-217, 240-243, 318-321; Ck2_Phospho_Site 52-55, 125-128, 195-198, 241-244, 242-245, 320-323, 325-328, 407-410, 412-415; Glycosaminoglycan 337-340; Myristyl 28-33, 72-77, 74-79, 80-85, 94-99, 100-105, 148-153, 152-157, 269-274, 420-425, 423-428; Pkc_Phospho_Site 45-47, 54-56, 163-165, 194-196, 409-411;	PS00152, SEA, SEA, SEA,
DEX0432_036.aa.16	y	2 - i7-29o177-199i	171-206,1.349; 114-124,1.225; 6-24,1.22; 149-169,1.158; 55-70,1.133; 93-102,1.088; 233-240,1.086; 130-136,1.082; 258-270,1.08; 104-110,1.058; 42-48,1.058; 84-90,1.047	Asn_Glycosylation 78-81, 156-159; Ck2_Phospho_Site 52-55, 79-82, 80-83, 158-161, 163-166, 245-248, 250-253; Glycosaminoglycan 175-178; Myristyl 28-33, 107-112, 258-263, 261-266; Pkc_Phospho_Site 45-47, 54-56, 247-249;	SEA, SEA, SEA,
DEX0432_036.aa.17	Y	0 -o	6-21,1.22; 326-344,1.216; 107-	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212,	PS00152, SEA, SEA, SEA,

			120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 158-170,1.088; 138-150,1.088; 304-313,1.088; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	263-266, 289-292; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294; Glycosaminoglycan 332-335; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323, 333-338; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	
DEX0432_036.aa.19	Y	0 -o	6-21,1.22; 107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 304-313,1.111; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 217-228,1.077; 124-130,1.065; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	PS00152, SEA,
DEX0432_036.aa.20	y	2 - i7-29o13 4-156i	128-163,1.349; 69-81,1.225; 6-24,1.22; 106-126,1.158; 190-197,1.086; 87-93,1.082; 215-227,1.08;	Asn_Glycosylation 113-116; Ck2_Phospho_Site 52-55, 67-70, 115-118, 120-123, 202-205, 207-210; Glycosaminoglycan 132-135; Myristyl 28-33, 215-220, 218-223;	SEA,

			42-48,1.058	Pkc_Phospho_Site 45-47, 54-56, 64-66, 67-69, 204-206;	
DEX0432_036.aa.21	y	1 -o518-540i	512-547,1.349; 409-431,1.227; 325-335,1.225; 6-21,1.22; 360-374,1.186; 107-120,1.149; 262-281,1.133; 456-475,1.133; 497-510,1.123; 177-190,1.12; 65-74,1.12; 23-33,1.114; 438-451,1.112; 245-257,1.112; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 304-313,1.088; 574-581,1.086; 341-347,1.082; 599-611,1.08; 217-228,1.077; 124-130,1.065; 381-392,1.058; 315-321,1.058; 51-57,1.058; 489-495,1.047; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292, 367-370, 457-460, 483-486; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372, 434-437, 484-487, 485-488, 586-589, 591-594; Glycosaminoglycan 516-519; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323, 393-398, 397-402, 403-408, 412-417, 422-427, 599-604, 602-607; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245, 588-590;	PS00152, SEA, SEA, SEA,
DEX0432_036.aa.22	Y	0 -o	316-342,1.228; 6-21,1.22; 107-120,1.149; 344-369,1.148; 262-	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292; Ck2_Phospho_Site 61-64, 134-137,	PS00152, PS00261, SEA,

			281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 217-228,1.077; 124-130,1.065; 51-57,1.058; 295-301,1.047	154-157, 174-177, 244-247, 290-293, 291-294; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 312-317, 321-326, 323-328; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245; Glyco_Hormone_Beta_1 328-334;	
DEX0432_036.aa.23	Y	0 -o	6-24,1.22; 98-111,1.149; 213-232,1.133; 128-141,1.12; 56-65,1.12; 196-208,1.112; 149-156,1.105; 77-85,1.092; 168-179,1.077; 115-121,1.065; 42-48,1.058; 246-252,1.047	Atpase_Alpha_Beta 138-147; Asn_Glycosylation 142-145, 160-163, 214-217, 240-243; Ck2_Phospho_Site 52-55, 125-128, 195-198, 241-244, 242-245; Myristyl 28-33, 72-77, 74-79, 80-85, 94-99, 100-105, 148-153, 152-157; Pkc_Phospho_Site 45-47, 54-56, 163-165, 194-196;	PS00152, SEA,
DEX0432_036.aa.24	y	1 - o388-410i	382-417,1.349; 325-335,1.225; 6-21,1.22; 360-380,1.158; 107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 158-170,1.088; 138-150,1.088; 304-313,1.088;	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292, 367-370; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372, 374-377, 456-459, 461-464; Glycosaminoglycan 386-389; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323, 487-492; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245, 458-460, 488-490,	PS00152, SEA, SEA, SEA,

			444-451,1.086; 341-347,1.082; 217-228,1.077; 458-464,1.073; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	491-493;	
DEX0432_036.aa.25	y	1 - o388-410i	382-417,1.349; 325-335,1.225; 6-21,1.22; 360-380,1.158; 107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 447-454,1.106; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 304-313,1.088; 341-347,1.082; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292, 367-370; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372, 374-377; Glycosaminoglycan 386-389; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245, 459-461;	PS00152, SEA, SEA, NLS_BP, SEA,
DEX0432_036.aa.26	Y	0 -o	6-24,1.22; 75-83,1.147; 110-117,1.086; 135-147,1.08; 42-48,1.058; 66-72,1.047	Asn_Glycosylation 60-63; Ck2_Phospho_Site 52-55, 61-64, 62-65, 122-125, 127-130; Myristyl 28-33, 135-140, 138-143; Pkc_Phospho_Site 45-47, 54-56,	SEA,

				124-126;	
DEX0432_036.aa.27	Y	0 -o	6-24,1.22; 93-101,1.147; 55-70,1.133; 128-135,1.086; 153-165,1.08; 42-48,1.058; 84-90,1.047	Asn_Glycosylation 78-81; Ck2_Phospho_Site 52-55, 79-82, 80-83, 140-143, 145-148; Myristyl 28-33, 153-158, 156-161; Pkc_Phospho_Site 45-47, 54-56, 142-144;	SEA,
DEX0432_036.aa.28	N	0 -o	4-10,1.093; 39-46,1.086; 64-76,1.08	Ck2_Phospho_Site 51-54, 56-59; Myristyl 64-69, 67-72; Pkc_Phospho_Site 53-55;	
DEX0432_036.aa.29	Y	0 -o	325-335,1.225; 6-21,1.22; 107-120,1.149; 262-281,1.133; 360-377,1.124; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 304-313,1.088; 341-347,1.082; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292, 367-370; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	PS00152, SEA, SEA, SEA,
DEX0432_036.aa.30	Y	0 -o	4-16,1.196; 20-26,1.11	Ck2_Phospho_Site 33-36; Myristyl 22-27;	
DEX0432_036.aa.31	y	1 -o168-190i	162-197,1.349; 105-115,1.225; 6-21,1.22; 140-160,1.158; 23-33,1.114;	Asn_Glycosylation 69-72, 147-150; Ck2_Phospho_Site 61-64, 70-73, 71-74, 149-152, 154-157, 236-239, 241-244;	SEA, SEA, SEA,

			84-93,1.088; 224-231,1.086; 121-127,1.082; 249-261,1.08; 95-101,1.058; 51-57,1.058; 75-81,1.047	Glycosaminoglycan 166-169; Myristyl 37-42, 98-103, 249-254, 252-257; Pkc_Phospho_Site 54-56, 63-65, 238-240;	
DEX0432_036.aa.32	y	1 - o145-167i	139-174,1.349; 82-92,1.225; 6-21,1.22; 117-137,1.158; 23-33,1.114; 63-70,1.095; 201-208,1.086; 98-104,1.082; 226-238,1.08; 72-78,1.058; 51-57,1.058	Asn_Glycosylation 124-127; Ck2_Phospho_Site 61-64, 126-129, 131-134, 213-216, 218-221; Glycosaminoglycan 143-146; Myristyl 37-42, 75-80, 226-231, 229-234; Pkc_Phospho_Site 54-56, 63-65, 215-217;	SEA, SEA, SEA,
DEX0432_036.aa.33	N	0 -o	4-17,1.195; 44-51,1.086; 58-64,1.073	Ck2_Phospho_Site 56-59, 61-64; Myristyl 87-92; Pkc_Phospho_Site 58-60, 88-90, 91-93;	
DEX0432_036.aa.34	N	1 - o324-346i	318-353,1.349; 261-271,1.225; 296-316,1.158; 43-56,1.149; 198-217,1.133; 113-126,1.12; 4-10,1.12; 181-193,1.112; 134-141,1.105; 22-30,1.092; 74-86,1.088; 94-106,1.088; 240-249,1.088; 380-387,1.086; 277-283,1.082; 405-417,1.08; 153-164,1.077; 60-66,1.065; 251-257,1.058; 231-237,1.047	Atpase_Alpha_Beta 123-132; Asn_Glycosylation 127-130, 145-148, 199-202, 225-228, 303-306; Ck2_Phospho_Site 70-73, 90-93, 110-113, 180-183, 226-229, 227-230, 305-308, 310-313, 392-395, 397-400; Glycosaminoglycan 322-325; Myristyl 17-22, 19-24, 25-30, 39-44, 45-50, 133-138, 137-142, 254-259, 405-410, 408-413; Pkc_Phospho_Site 148-150, 179-181, 394-396;	PS00152, SEA, SEA, SEA,

DEX0432 _036.aa .35	y	1 - o389- 411i	383- 418,1.349; 326- 336,1.225; 7- 22,1.22; 361- 381,1.158; 108- 121,1.149; 263- 282,1.133; 178-191,1.12; 66-75,1.12; 24-34,1.114; 246- 258,1.112; 199- 206,1.105; 87-95,1.092; 159- 171,1.088; 139- 151,1.088; 305- 314,1.088; 445- 452,1.086; 342- 348,1.082; 470-482,1.08; 218- 229,1.077; 125- 131,1.065; 316- 322,1.058; 52-58,1.058; 296-302,1.047	Atpase_Alpha_Beta 188-197; Asn_Glycosylation 192-195, 210-213, 264-267, 290-293, 368-371; Ck2_Phospho_Site 62-65, 135-138, 155-158, 175-178, 245-248, 291-294, 292-295, 370-373, 375-378, 457-460, 462-465; Glycosaminoglycan 387-390; Myristyl 38-43, 82-87, 84- 89, 90-95, 104- 109, 110-115, 198-203, 202-207, 319-324, 470-475, 473-478; Pkc_Phospho_Site 55-57, 64-66, 213-215, 244-246, 459-461;	PS00152, SEA, SEA, SEA,
DEX0432 _036.aa .36	y	2 - i7- 29o15 9- 181i	153- 188,1.349; 96-106,1.225; 6-24,1.22; 131- 151,1.158; 75-84,1.088; 215- 222,1.086; 112- 118,1.082; 240-252,1.08; 86-92,1.058; 42-48,1.058; 66-72,1.047	Asn_Glycosylation 60-63, 138-141; Ck2_Phospho_Site 52-55, 61-64, 62- 65, 140-143, 145- 148, 227-230, 232-235; Glycosaminoglycan 157-160; Myristyl 28-33, 89-94, 240-245, 243-248; Pkc_Phospho_Site 45-47, 54-56, 229-231;	SEA, SEA, SEA,
DEX0432 _036.aa .37	y	1 - o421- 443i	415- 450,1.349; 358- 368,1.225; 6- 21,1.22; 320- 346,1.163; 393- 413,1.158;	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292, 400-403; Ck2_Phospho_Site 61-64, 134-137,	PS00152, SEA, SEA, SEA,

			107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 304-318,1.111; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 477-484,1.086; 374-380,1.082; 502-514,1.08; 217-228,1.077; 124-130,1.065; 348-354,1.058; 51-57,1.058; 295-301,1.047	154-157, 174-177, 244-247, 290-293, 291-294, 319-322, 402-405, 407-410, 489-492, 494-497; Glycosaminoglycan 419-422; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 351-356, 502-507, 505-510; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245, 491-493;	
DEX0432_036.aa.38	N	1 - o84-106i	78-113,1.349; 21-31,1.225; 56-76,1.158; 140-147,1.086; 37-43,1.082; 165-177,1.08; 11-17,1.058	Asn_Glycosylation 63-66; Ck2_Phospho_Site 65-68, 70-73, 152-155, 157-160; Glycosaminoglycan 82-85; Myristyl 14-19, 165-170, 168-173; Pkc_Phospho_Site 154-156;	SEA, SEA, SEA,
DEX0432_036.aa.39	y	2 - i7-29o136-158i	130-165,1.349; 73-83,1.225; 6-24,1.22; 108-128,1.158; 54-61,1.095; 192-199,1.086; 89-95,1.082; 217-229,1.08; 63-69,1.058; 42-48,1.058	Asn_Glycosylation 115-118; Ck2_Phospho_Site 52-55, 117-120, 122-125, 204-207, 209-212; Glycosaminoglycan 134-137; Myristyl 28-33, 66-71, 217-222, 220-225; Pkc_Phospho_Site 45-47, 54-56, 206-208;	SEA, SEA, SEA,
DEX0432_036.aa.41	Y	0 -o	325-335,1.225; 6-21,1.22; 373-393,1.155; 107-	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292,	PS00152, SEA, SEA, SEA,

			120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 158-170,1.088; 138-150,1.088; 304-313,1.088; 360-369,1.083; 341-347,1.082; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	367-370; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	
DEX0432_036.aa.42	Y	0 -o	6-21,1.22; 107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 304-313,1.088; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	PS00152, SEA,
DEX0432_036.aa.43	Y	0 -o	325-335,1.225; 6-21,1.22; 360-379,1.189;	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212,	PS00152, SEA, SEA, SEA,

			107- 120,1.149; 381- 406,1.148; 262- 281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245- 257,1.112; 198- 205,1.105; 86-94,1.092; 138- 150,1.088; 158- 170,1.088; 304- 313,1.088; 341- 347,1.082; 217- 228,1.077; 124- 130,1.065; 315- 321,1.058; 51-57,1.058; 295-301,1.047	263-266, 289-292, 367-370; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 369-372; Myristyl 37-42, 81-86, 83-88, 89- 94, 103-108, 109- 114, 197-202, 201-206, 318-323; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	
DEX0432 _036.aa .44	Y	0 -o	6-21,1.22; 107- 120,1.149; 304- 312,1.147; 262- 281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245- 257,1.112; 198- 205,1.105; 86-94,1.092; 138- 150,1.088; 158- 170,1.088; 339- 346,1.086; 364-376,1.08; 217- 228,1.077; 124- 130,1.065; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294, 351-354, 356-359; Myristyl 37-42, 81-86, 83- 88, 89-94, 103- 108, 109-114, 197-202, 201-206, 364-369, 367-372; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245, 353-355;	PS00152, SEA,
DEX0432 _036.aa	y	1 - o15-	8-43,1.349; 70-77,1.086;	Ck2_Phospho_Site 82-85, 87-90;	

.45		37i	95-107,1.08	Myristyl 95-100, 98-103; Pkc_Phospho_Site 84-86;	
DEX0432 _036.aa .46	Y	0 -o	6-24,1.22; 53-60,1.175; 87-94,1.086; 112-124,1.08; 42-48,1.058	Ck2_Phospho_Site 52-55, 99-102, 104-107; Myristyl 28-33, 112-117, 115-120; Pkc_Phospho_Site 45-47, 54-56, 101-103;	
DEX0432 _036.aa .48	Y	0 -o	6-21,1.22; 107-120,1.149; 262-281,1.133; 177-190,1.12; 65-74,1.12; 23-33,1.114; 245-257,1.112; 198-205,1.105; 86-94,1.092; 138-150,1.088; 158-170,1.088; 304-313,1.088; 217-228,1.077; 124-130,1.065; 315-321,1.058; 51-57,1.058; 295-301,1.047	Atpase_Alpha_Beta 187-196; Asn_Glycosylation 191-194, 209-212, 263-266, 289-292; Camp_Phospho_Site 337-340; Ck2_Phospho_Site 61-64, 134-137, 154-157, 174-177, 244-247, 290-293, 291-294; Myristyl 37-42, 81-86, 83-88, 89-94, 103-108, 109-114, 197-202, 201-206, 318-323; Pkc_Phospho_Site 54-56, 63-65, 212-214, 243-245;	PS00152, SEA,
DEX0432 _039.aa .1				Myristyl 25-30;	
DEX0432 _041.aa .1				Ck2_Phospho_Site 79-82; Myristyl 35-40;39-44;61-66;	
DEX0432 _045.aa .1				Asn_Glycosylation 45-48; Glycosaminoglycan 35-38; Myristyl 29-34;	

Using the PSORT II program, the following cellular localizations and the k nearest neighbors classifier values were determined (Paul Horton and Kenta Nakai, Better Prediction of Protein Cellular Localization Sites with the k Nearest Neighbors Classifier, Intelligent Systems for Molecular Biology 5 147-152 (1997).

DEX ID NO	Localization	K value
DEX0432_5.aa.1	nuc	(k=23)
DEX0432_10.aa.1	nuc	(k=23)
DEX0432_12.aa.1	cyt	(k=23)
DEX0432_13.aa.1	nuc	(k=23)
DEX0432_15.aa.1	nuc	(k=23)
DEX0432_16.aa.1	pla	(k=23)
DEX0432_17.aa.1	cyt	(k=23)
DEX0432_18.aa.1	nuc	(k=23)
DEX0432_19.aa.1	nuc	(k=23)
DEX0432_21.aa.1	cyt	(k=23)
DEX0432_23.aa.1	cyt	(k=23)
DEX0432_25.aa.1	cyt	(k=23)
DEX0432_26.aa.1	ves	(k=9)
DEX0432_31.aa.1	pla	(k=23)
DEX0432_33.aa.1	nuc	(k=23)
DEX0432_35.aa.1	nuc	(k=23)
DEX0432_39.aa.1	exc	(k=9)
DEX0432_41.aa.1	nuc	(k=23)
DEX0432_45.aa.1	nuc	(k=23)

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1-94. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991).

Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-

release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in

the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA

ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-In Vivo

Another aspect of the present invention is using in vivo gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318,

Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, and Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290.

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, breast, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, breast, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are

differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to breasts or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for

protein expression may be done in a similar fashion except that quadriceps from different mïce are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (I. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989). For a review of such techniques, see Gordon,"Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989).

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in

order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or"knocking out"the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989)). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered

not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and Mulligan & Wilson, U. S. Patent No. 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and"knock-out"animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

We claim:

1. An isolated nucleic acid molecule comprising:

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 95-156;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is an RNA.

5. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

6. The nucleic acid molecule according to claim 5, wherein the nucleic acid molecule is a human nucleic acid molecule.

7. A method for determining the presence of a breast specific nucleic acid (BSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule of SEQ ID NO: 1-94 under conditions in which the nucleic acid molecule will selectively hybridize to a breast specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a BSNA in the sample, wherein the detection of the hybridization indicates the presence of a BSNA in the sample.

8. A vector comprising the nucleic acid molecule of claim 1.

9. A host cell comprising the vector according to claim 8.

10. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of:

(a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and

(b) incubating the host cell under conditions in which the polypeptide is produced.

11. A polypeptide encoded by the nucleic acid molecule according to claim 1.

12. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 95-156 ; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94.

13. An antibody or fragment thereof that specifically binds to:

(a) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 95-156 ; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94.

14. A method for determining the presence of a breast specific protein in a sample, comprising the steps of:

(a) contacting the sample with a suitable reagent under conditions in which the reagent will selectively interact with the breast specific protein comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 95-156; and

(b) detecting the interaction of the reagent with a breast specific protein in the sample, wherein the detection of binding indicates the presence of a breast specific protein in the sample.

15. A method for diagnosing or monitoring the presence and metastases of breast cancer in a patient, comprising the steps of:

(a) determining an amount of:

(i) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 95-156;

(ii) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94;

(iii) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (i) or (ii);

(iv) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (i) or (ii);

(v) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 95-156 ; or

(vi) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94 and;

(b) comparing the determined amount of the nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the breast specific marker in a normal control; wherein a difference in the determined amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of

the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of breast cancer.

16. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence of:

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 95-156;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b); or

(e) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 95-156 ; or

(f) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94.

17. A method of treating a patient with breast cancer, comprising the step of administering a composition consisting of:

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 95-156;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b);

(d) a nucleic acid molecule having at least 95% sequence identity to the nucleic acid molecule of (a) or (b);

(e) a polypeptide comprising an amino acid sequence with at least 95% sequence identity to of SEQ ID NO: 95-156 ; or

(f) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 95% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-94;

to a patient in need thereof, wherein said administration induces an immune response against the breast cancer cell expressing the nucleic acid molecule or polypeptide.

18. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 12.

SEQUENCE LISTING

<110> diaDexus, Inc.
Salceda, Susana
Macina, Roberto A.
Turner, Leah R.
Sun, Yongming
Liu, Chenghua

<120> Compositions and Methods Relating to Breast Specific Genes and Proteins

<130> DEX-0432

<150> US 60/389,327
<151> 2002-06-14

<160> 171

<170> PatentIn version 3.1

<210> 1
<211> 1574
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (89)..(180)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (1466)..(1466)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (1474)..(1474)
<223> n=a, c, g, or t

<400> 1
ctgaaggttt atacaatatt tacacagtgg ctacaatatt cacaaaattc ttatgttctc 60

ttatgaaaaa tatacacttt tcattttgnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180

gtacatgtat atatttgtcc tgcattatgt tttttacttg atataaatgt atttttactg 240

tgatagtcca agtgccctgg ggggcaggtg tgctctatgt ggttcttctt ccattggaga 300

gctggcgtag agatctgcag tgttcacaag gatgttggtt tggagatgtc tgctgctagg 360

acctggggtg tgtgactcag tccatatgag agggacatct gggtggagga gtaaattcct 420

gtgctctgaa atgccacttg gtagctctgg acaatgaagg acaattgact caagggtgcc 480

tggcttctgc tgctgctggg aaaaaattca gtttatagca ttcctgcacc tcccaaagta 540

gataacctgg aggtcattca gttaacaact gtccctgagg actcagtttt gggggagggg 600

ttatctggga gaagctttag cctgttctga gccattagga gacattagtg aattggagca 660

ctggagaatc ctacaaatgg cctatgtctc agaagagctg ggacctcctt ccagctgctg 720

cagatgctga caggccctgg gaggctgctg tgctctggag aagctggagc agctcatttc 780

ttggcctagc ctggctgcct cagaaagagc agtcaggact tgagggaagc atcaaattct 840

atacccataa actgcagttg gaagtcagct ttttgaaatg tccagccttt gcccaattgt 900

ttcagatcat ctcatgcctc aggctttggc aggtatcctg ccctccatct tattccagtg 960

tgttcacctc atcaaggcag cagagtggat gaaggagtaa gtctgccctt tgccatactg 1020

aacagctgtg gaccccgatt ggtgagggct ctgcatatgc ctgtatgaag gagatacagg 1080

tgtgtgtgca catgccggta tgaagaagac acaggcatgt gcttctcagt tttgctaaca 1140

gtgggagctc aacggggcag agggaggaag gtccatgatg ctcagccaca tactgtagag 1200

agaggcaatt taatgttaaa tgacgcacca tcctccctcc caccottctc ccagtcaact 1260

ttttttcttt ttctagaact actaattatc tctcaaggct gaaaaattaa ttgccttagg 1320

tggagaactt aattcctagt atccaccaaa cttaactccg tatctccata tggtgtctcc 1380

atatctactg tgtgagctac ttaactgacg ccctcttcct ccaactgaag gatcgcccaa 1440

cgtttttgga ttatagaatt attatngcct gctntctttc tttgggactt ttgaatttct 1500

ttggtttcgt ttttaagaag taacccaaca tttcctacaa cactaaataa aatggtactt 1560

acctttcaaa aaga 1574

<210> 2
<211> 539
<212> DNA
<213> Homo sapien

<400> 2
cgaccgttga ctattctcta caaaccacaa agacattgga acactatacc tattattcgg 60

cgcatgagct ggagtcctag gcacagctct aagcctcctt attcgagccg agctgggcca 120

gccaggcaac cttctaggta acgaccacat ctacaacgtt atcgtcacag cccatgcatt 180

tgtaataatc ttcttcatag taatacccat cataatcgga ggctttggca aaaaaaaaac 240

aaaaaaaaaa aaaaaacctg gggaaacac ggggcaaacg cggtcccggg ggcagaaatg 300

gtacccggcc acattcccac acacattccg acacaagagg cgaagacacg acaacagccg 360

accgacacaa cagaggcacg gggaaggggg acgaagagga ggaggagaac agacgggacg 420

gcaacaaggg acagcgaggg acgcagacgc ggaggagaag gggaaggca gacgggaacg 480

agaaaaagag ccgagacggg acgcggaccc cacagggggg tcgcgagaaa agacgccca 539

<210> 3
<211> 197
<212> DNA
<213> Homo sapien

<400> 3
actttttatt caatgtaatc agaagctgtg atgttttgcc tttgtagtcc tgtgctttgt 60

tactgtaatt tttttttttt ttatacaaag cacgtgacgt ggactaatgt aaggcagatg 120

acgtgactct taagacgtgc tatatttatt cagttcctct ttacctctat agaggtttta 180

aatttagaat aagctgt 197

<210> 4
<211> 1634
<212> DNA
<213> Homo sapien

<400> 4
aactaacttt gtggggtttt tttgtttttg tttttatttt cttaaagccg aacgagcatg 60

tatgtggaca gaagtggaac acttcttggg tccaaataca agaaagtctt atatcgtcaa 120

tatgatgata acacgttcac aaatcaaaca aaaaggaatg aaggtgaaaa acatctcgat 180

atactaggtc cattaatatt gctcaaccct ggtcaaataa ttcaaattat ctttaaaaat 240

aaagccgcaa gaccgtattc tattcatgct catggagtga aaacaataa ttccactgtt 300

gttccaactc agccaggaga gattcaaata tatacttggc agatacctga tagaactggt 360

cctacctcac tggactttga atgcatacct tggttttact attcaactgt atctgtggct 420

aaggaccttc acagtggact ggtaggccct ctctctgtat gccgcaaaga catcaacccc 480

aacatagttc accgtgttct ccacttcatg atatttgatg agaatgaatc ctggtacttc 540

gaagacagta tcaacaccta tgcttcaaaa ccaaacaaag tggacaagga aaatgataat 600

tttcaactca gcaaccaaat gcacgcaatt aacggaagac tgtttggaaa taaccaaggt 660

ataacattcc atgttgggga tgtagtgaat tggtatctga ttggcatagg gaatgaagct 720

gacctgcaca cagttcactt tcatggccat agctttgaat acaagaatta gggagtgtat 780

caatctgatg tttatgacct tcctcctggg gtctatcgaa ctgtaaaaat gtatcgaaga 840

gatgttggaa cctggttatt ttattgccat gtttttgagc acattggtgc tggaatggat 900

agcacttaca ctgtacttga aagaaaaggg ctgatggagc agaacctctg aagcagacaa 960

aggagagtca gcatgaacag tttctcagaa tcttctctca atatcaggac tacatttgtc 1020

aacaaaacca aaaactgatt agccaccgat ataattttta cctacaacat cctattaatg 1080

tcaataatat cattattgat acaattctaa taatcactac ccttattcct atcagtgttc 1140

atgtacattc ttagtaaaag agactttggt gcgctgtcca tgaaataaat cccccattgc 1200

taacattctt tctttggaaa agtagatttt gcatttcaaa gaatataaag tcaaattgga 1260

ttggatttac aggtcatctg ttcccacaga agggtgatat tgatgttgct attgataagt 1320

aaactttttg tggcaaaagt gatggtagtt attttaagga tgttcccaag actaatataa 1380

attttgtatt tattccttaa atgtatgtaa tcattttagc ttagtatttt aacttagaac 1440

tgcatgctat tatataatat tacctatttt tgaaacttcc ttttctacag cataaatatt 1500

tgatatgata tgaatattga caagcttaca agccaaggta aagctgccaa agaaggaaaa 1560

ctccagggac caaggagtct gggaggaacc agctaaagac tttcatgaca atgtaccagg 1620

gagactagtt tgag 1634

<210> 5
<211> 891
<212> DNA
<213> Homo sapien

<400> 5
ggggaagtgc aggatggggg gacaggggcg ctcccgcggg ggtggatgag ggaccatagc 60

ggggctggcg gggcaggggc cggcgcacga ggctggagga ggggagcgcg cgcttctacc 120

cgggctgggt cgccgagtcc acagcctcga agccatgggt tctccccggc cctctgaagc 180

cgccacacct gtgccagccg gccgcgtcct cagacctttc cccgcggagt cttcccagca 240

cttggagacg cagcgcaggg cccggaggac ggcctggccc ggagaaaaga taccgaagct 300

ccaactttcc ccaaccccgc tcccctcctc cttccaccct cccttcccgc ccccaaagct 360

cgggggtcct atccctcctc cggtccgcgg agtctcccga accctgcggg gacccggcgc 420

tcggcggtgc cctcctgggg cgcacggggc tggggcggga gcgaggagac caggtgggga 480

ggggacccca gatctcagac gccaggggag acggcgtttc ccgctgttca ttcaggtttg 540

tgccaaaagg agcctcacag atgcagtatt gggtttggta gactcaaatc gtcttgtttt 600

aatgtaaatg aaagtaagtt taggataaat tccagtgcgg cgggggcagg caaggctacc 660

cacatttttt aaaaagaagc cagcccgtat ttttctccct ttccaaatcc tccgcccccc 720

agtccttcga cccaggcacg agcgcccatc gcggaggcca cgatgcccgt tttattccct 780

ctccacggca aggaaaagca gcgaaatctg aggtcttcag aggttaaccc tatctaggag 840

cagaatgtga cgcattgtaa acaaataaat attgaaaact cgatgttaaa a 891

<210> 6
<211> 1253
<212> DNA
<213> Homo sapien

<400> 6
```
ggggaagtgc aggatggggg gacaggggcg ctcccgcggg ggtggatgag ggaccatagc      60
ggggctggcg gggcaggggc cggcgcacga ggctggagga ggggagcgcg cgcttctacc     120
cgggctgggt cgccgagtcc acagcctcga agccatgggt tctccccggc cctctgaagc     180
cgccacacct gtgccagccg gccgcgtcct cagacctttc cccgcggagt cttcccagca     240
cttggagacg cagcgcaggg cccggaggac ggcctggccc ggagaaaaga taccgaagct     300
ccaactttcc ccaaccccgc tcccctcctc cttccaccct cccttcccgc ccccaaagct     360
cgggggtcct atccctcctc cggtccgcgg agtctcccga accctgcggg gacccggcgc     420
tcggcggtgc cctcctgggg cgcacggggc tggggcggga gcgaggagac caggtgggga     480
ggggacccca gatctcagac gccaggggag acggcgtttc ccgctgttca ttcaggtttg     540
tgccaaaagg agcctcacag atgcagtatt gggtttggta gactcaaatc gtcttgtttt     600
aatgtaaatg aaagtaagtt taggataaat tccagtgcgg cgggggcagg caaggctacc     660
cacatttttt aaaaagaagc cagcccgtat ttttctccct ttccaaatcc tccgcccccc     720
agtccttcga cccaggcacg agcgcccatc gcggaggcca cgatgcccgt tttattccct     780
ctccacggca aggaaaagca gcgaaatctg aggtcttcag aggttaaccc tatctaggag     840
cagaatgtga cgcattgtaa acaaataaat attgaaaact cgatgttaaa ccctttactt     900
tttctgactc cgacttgctt gacctctgag cagacctggg tttcgaacac agacgccctt     960
ccccatttct ctattctctg tattcctgtt tcaccttcac agcagtctgc cagcacttct    1020
tagcactcag tttaaccaga gcacaagctc ctgaatagca aaaaccaggt ctttttatac    1080
gtggcacagt ggctgttaca aaatatgctt cttgggtgaa ttggtaaaaa atattgtatt    1140
actttttatt tgtagcaaaa cctagaataa gaaaaagtac aagagattat tgtttgcctt    1200
taaattgcat ttttaaaaga gcgtgcatat aatctctgag aaattaaatg tct            1253
```

<210> 7
<211> 401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (144)..(144)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (174)..(174)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (304)..(304)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (383)..(384)
<223> n=a, c, g, or t

<400> 7
acgttcaaag caggcgaact tcatcatggt gtatggtatc tgtctcatcc agagaggagc 60

aaccccctat gtagaatgct tttagagcct tcttcctata tacatttctg ggagctgcat 120

ccactcaaag tgcttggcat aacnctggct ggcgtttgca attacagaac cttnacgcag 180

cttccactag gcacgccagg agcaagtgtc acgcacaaga cattttcagc actggcagac 240

ggcatgccaa catatacgtg catgctcgcg ccagagcata cagtattccc tcctaaagat 300

ccanacaaca caaggcaagg gcatgctgca attgcctgtt ggtgttaggt ctttcacatt 360

cgacatgtga acagttctta gannacaaca acttaagctt g 401

<210> 8
<211> 405
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (56)..(57)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (69)..(70)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (77)..(77)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (79)..(80)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (102)..(102)

<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (200)..(200)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (247)..(247)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (250)..(251)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (274)..(275)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (286)..(287)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (295)..(295)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (297)..(298)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (306)..(306)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (309)..(309)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (317)..(318)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (337)..(337)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (339)..(340)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (347)..(347)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (349)..(350)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (356)..(357)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (374)..(375)
<223> n=a, c, g, or t

<400> 8
actatttaaa atgctcaatt tcagcaccga tggccatgta aataagatga tttaanntgt 60

tgattttann atcctgntnn atataaaata acaaagtcac anatgagttt gggcatattt 120

aatgatgatt atggagcctt agaggtcttt aatcattggt tcggctgctt ttatgtagtt 180

taggctggaa atggtttcan cttgctcttt gacgtgtcac gcaagactga acgatagctt 240

ttcctgngan ncagctagaa aacacaagaa tctnntgtag gtacannttg caccntnnat 300

ctcagncgnc ataggtnngc agtcttcgct tctacantnn gatgctnann aaggcnntgc 360

gaactgcgga ctcnnctgat gcgacactaa ggactccaat gtcga 405

<210> 9
<211> 305
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature

<222> (1)..(19)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (286)..(305)
<223> n=a, c, g, or t

<400> 9
nnnnnnnnnn nnnnnnnnnt aaaaagaaaa aaggaaactg gttacacatc tgtccacaaa 60

ggcaaatgca ggggggctgg tgactcctgg gtataaaggc tcacatctgt ttatgttaat 120

taagagagca gtatgtaacc agtatcattc cacttcagtt ttcttttagg atctaacata 180

gtgctatcca agagatatat aatataatgc cacatgttat atttcctgat agcctcattt 240

tataaagtag tccaatgctt cactcagcca ttttacctca cccccnnnnn nnnnnnnnnn 300

nnnnn 305

<210> 10
<211> 299
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (280)..(299)
<223> n=a, c, g, or t

<400> 10
ggagtgtgct tgtggcccct cagaaagata gtctgctggc tcctaggggt tggggtgggg 60

gacacacctt tttctcagga agaggtgatg gcaatgtaaa acatctaagc aaagttttaa 120

atgaaaaaaa ggaaacacat ttaaacatcc tgataatgga gggaaggggg gcacatttac 180

acatagccca gaacttgtag aattctgcat agtgaatgta tattgaatta gtctcctgcc 240

ttatacattc aggaggaata aatttccata atgtaaggcn nnnnnnnnnn nnnnnnnnn 299

<210> 11
<211> 1249
<212> DNA
<213> Homo sapien

<400> 11
tagctccttc caactcctca gaatctccac tctatggatc tggacctctg gattcggctt 60

tctccctggg cactgccttc aggaagacgt tgagaattga ccttacacaa tcccagcgcc 120

ctcctcacag gagcctttca ctttacagtg gcaaggggcc tggttctgga gaactggctg 180

atgctctgaa tttcttcata taccccacat ttgactttgg cttacactgt acaattggag 240

atgttgctac aggtccctga gatgcaatca gattaagcgt agcaagcatt gccaatggga 300

aagtcaaaat aatttatttt ttttcccttt ccccctaccc catccccagc caagaatttc 360

ttttcaagat atcgtcatca ttcttaaaca acattcttaa cccccagctg gggtccccat 420

tttaatagat gtcattgctt caagtctaac ggcgccggga ggcctgtttg agggaaaaca 480

ttagtttgaa aaatccccgt tcccttcatc cactgccctt gttctccacg tgggagtgtg 540

cttgtggccc ctcagaaaga tagtctgctg gctcctaggg gttggggtgg gggacacacc 600

tttttctcag gaagaggtga tggcaatgta aaacatctaa gcaaagtttt aaatgaaaaa 660

aaggaaacac atttaaacat cctgataatg gagggaaggg gggcacattt acacatagcc 720

cagaacttgt agaattctgc atagtgaatg tatattgaat tagtctcctg ccttatacat 780

tcaggaggaa taaatttcca taatgtaagg caaatgcatg gggttctgag gttcactttg 840

caagtgccct tgctgccttt cctctgtgtc tattatggct ctttaagttg acggttcctg 900

gagcagcttg tatttagttt cgtttggcag tctggccctg ttgactttga tttgcagacc 960

aattctccct tgacctgact cacagccgcc tgctcttacc cccctcctca ggaagtcttc 1020

ctcattaaag gatgtgatga cggagctcag ggatgagaat gcacatgtga gactgtgtga 1080

caccaaggag ggttgtgcga actggtgaca acatggcagc accatggcct gtgggggttg 1140

tgtgactagt gtgactgtgc tggcgaccat atggacctgt tttgtcagtc ggtgtctaag 1200

caggagatgg cacactcaaa ctgggaagtg ttttaaacat aggctattc 1249

<210> 12
<211> 236
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (217)..(236)
<223> n=a, c, g, or t

<400> 12
tggccagaat cccccagaga atcagggacc agctttactg gagttggggg cggcttgtct 60

tcgctggctc ctaccccatc tccaagataa gcctgagcct tagctcccag ctagggggcg 120

ttatttatgg accactttta tttattgtca gacacttatt tattgggatg tgagccccag 180

gggggcctcc tcctaggata ataaacaatt ttgcggnnnn nnnnnnnnnn nnnnnn 236

<210> 13
<211> 3218
<212> DNA
<213> Homo sapien

<400> 13

cccgggcaaa agcgagcgcc gcccctgcct ctccgctgct ggctggaacg ctgatctatc 60

tagttgctgg ggagacgccc ccagatgccc gggccccact cggacttcag cacacatccc 120

gaaggatggg gaaagaaaga ggcccccacg agcgggactc gcagtggcca aggaggggtg 180

agaggcggac agggatcagc tggcccctgc ggcctggttg cacctgcatg gtgactagct 240

gccgggctgc gccccggggc gcggcgagga ggcggggtct ggcagtgcgt tgggtggggg 300

aggagcttct gggtgatgta aggccgggaa tgggagtggg cctctcctcg actcgctgct 360

aggaaggggg cgggactctc ggtgaccaga cgccggggag ggggcaggcg ttcattgata 420

aaacgctggg ctcccctggg cgccagcgca gcgtagcaaa tccaggcagc gccacgcgcg 480

gccggggccg ggcggaaccg agaagccggg accgcgctgc gacgcgccgg ccgcatggag 540

cctgccgccg gtttcctgtc tccgcgcccc ttccagcgtg cggccgccgc gcccgctccc 600

ccggccgggc ccgggccgcc tccgagtgcc ttgcgcggac ctgagctgga gatgctggcc 660

gggctaccga cgtcagaccc cgggcgcctc atcacggacc cgcgcagcgg ccgcacctac 720

ctcaaaggcc gcttgttggg caagggggc ttcgcccgct gctacgaggc cactgacaca 780

gagactggca gcgcctacgc tgtcaaagtc atcccgcaga gccgcgtcgc caagccgcat 840

cagcgcgaga agatcctaaa tgagattgag ctgcaccgag acctgcagca ccgccacatc 900

gtgcgttttt cgcaccactt tgaggacgct gacaacatct acattttctt ggagctctgc 960

agccgaaagt ccctggccca catctggaag gcccggcaca ccctgttgga gccagaagtg 1020

cgctactacc tgcggcagat cctttctggc ctcaagtact tgcaccagcg cggcatcttg 1080

caccgggacc tcaagttggg aaatttttc atcactgaga acatggaact gaaggtgggg 1140

gattttgggc tggcagcccg gttggagcct ccggagcaga ggaagaagac catctgtggc 1200

acccccaact atgtggctcc agaagtgctg ctgagacagg gccacggccc tgaggcggat 1260

gtatggtcac tgggctgtgt catgtacacg ctgctctgcg ggagccctcc ctttgagacg 1320

gctgacctga aggagacgta ccgctgcatc aagcaggttc actacacgct gcctgccagc 1380

ctctcactgc ctgcccggca gctcctggcc gccatccttc gggcctcacc ccgagaccgc 1440

ccctctattg accagatcct gcgccatgac ttctttacca agggctacac ccccgatcga 1500

ctccctatca gcagctgcgt gacagtccca gacctgacac cccccaaccc agctaggagt 1560

ctgtttgcca aagttaccaa gagcctcttt ggcagaaaga agaagagtaa gaatcatgcc 1620

caggagaggg atgaggtctc cggtttggtg agcggcctca tgcgcacatc cgttggccat 1680

caggatgcca ggccagaggc tccagcagct tctggcccag cccctgtcag cctggtagag 1740

acagcacctg aagacagctc accccgtggg acactggcaa gcagtggaga tggatttgaa 1800

gaaggtctga ctgtggccac agtagtggag tcagcccttt gtgctctgag aaattgtata 1860

gccttcatgc ccccagcgga acagaacccg gccccccctgg cccagccaga gcctctggtg 1920

tgggtcagca agtgggttga ctactccaat aagttcggct ttgggtatca actgtccagc 1980

cgccgtgtgg ctgtgctctt caacgatggc acacatatgg ccctgtcggc caacagaaag 2040

actgtgcact acaatcccac cagcacaaag cacttctcct tctccgtggg tgctgtgccc 2100

cgggccctgc agcctcagct gggtatcctg cggtacttcg cctcctacat ggagcagcac 2160

ctcatgaagg gtggagatct gcccagtgtg gaagaggtag aggtacctgc tccgcccttg 2220

ctgctgcagt gggtcaagac ggatcaggct ctcctcatgc tgtttagtga tggcactgtc 2280

caggtgaact tctacgggga ccacaccaag ctgattctca gtggctggga gcccctcctt 2340

gtgacttttg tggcccgaaa tcgtagtgct tgtacttacc tcgcttccca ccttcggcag 2400

ctgggctgct ctccagacct gcggcagcga ctccgctatg ctctgcgcct gctccgggac 2460

cgcagcccag cctaggaccc aagccctgag gcctgaggcc tgtgcctgtc aggctctggc 2520

ccttgccttt gtggccttcc cccttccttt ggtgcctcac tgggggcttt gggccgaatc 2580

ccccagggaa tcagggacca gctttactgg agttgggggc ggcttgtctt cgctggctcc 2640

taccccatct ccaagataag cctgagcctt agctcccagc taggggggcgt tatttatgga 2700

ccacttttat ttattgtcag acacttattt attgggatgt gagccccagg ggggcctcct 2760

cctaggataa taaacaattt tgcagaaaaa aaaaacaaca aaacaaaaaa acaaaacaga 2820

agcacacaac caccacacaa cacgaggggc cccaaccaag agaccaaccc acaaccgagc 2880

ccacaaacag agggacgcga cacaccgcac acgacacagg caagagcggg cgcacccaca 2940

acggaccgcc cgccacgggc agaggcagcg agggacgcac agatacacag aggaggaggc 3000

gagagaaaag ggaggagagg agagaacaac agaggagggc gaacacgacg cccgcggaga 3060

caagcgaggg cggccacacc caccaagagg agaccggaca acccgggaga aaacaaccgc 3120

gacagcgaca ggagggcgcc agagaggcag acacagagcg cagcgcggca cagagcgccg 3180

cgggagccgc cggacgacca gtacaacagg aacagcaa 3218

<210> 14
<211> 501
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (84)..(84)
<223> n=a, c, g, or t

<220>

<221> misc_feature
<222> (137)..(137)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (146)..(147)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (160)..(161)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (169)..(170)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (181)..(181)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (183)..(184)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (195)..(196)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (205)..(206)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (211)..(212)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (219)..(221)
<223> n=a, c, g, or t

<220>
<221> misc_feature

<222> (227)..(228)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (234)..(234)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (236)..(236)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (238)..(238)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (241)..(243)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (249)..(249)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (252)..(253)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (256)..(256)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (259)..(259)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (261)..(262)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (267)..(267)

<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (271)..(271)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (273)..(273)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (275)..(275)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (278)..(278)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (280)..(281)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (284)..(284)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (287)..(287)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (289)..(289)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (291)..(292)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (296)..(297)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (303)..(303)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (305)..(305)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (308)..(308)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (310)..(310)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (312)..(312)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (315)..(315)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (317)..(318)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (322)..(322)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (324)..(324)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (327)..(329)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (336)..(336)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (338)..(339)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (344)..(345)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (349)..(349)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (351)..(351)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (353)..(353)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (356)..(356)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (358)..(358)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (360)..(360)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (362)..(362)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (367)..(367)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (369)..(369)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (371)..(371)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (373)..(373)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (375)..(375)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (377)..(377)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (380)..(380)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (382)..(383)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (389)..(389)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (391)..(391)
<223> n=a, c, g, or t

<220>

<221> misc_feature
<222> (393)..(393)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (396)..(396)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (398)..(398)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (402)..(402)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (404)..(404)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (406)..(406)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (408)..(408)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (410)..(410)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (412)..(412)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (414)..(414)
<223> n=a, c, g, or t

<220>
<221> misc_feature

<222> (416)..(416)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (421)..(421)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (437)..(437)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (447)..(447)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (453)..(453)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (455)..(455)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (459)..(459)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (461)..(461)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (463)..(463)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (467)..(468)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (471)..(471)

<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (473)..(473)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (475)..(475)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (477)..(477)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (479)..(479)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (481)..(481)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (483)..(483)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (485)..(485)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (487)..(487)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (489)..(489)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (491)..(491)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (493)..(493)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (497)..(497)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (499)..(499)
<223> n=a, c, g, or t

<400> 14
acaggccgac agagaagatt cccgagagta aatcatcttt ccaatccaga ggaacaagca 60

tgtctctctg cgcaagatcc atcntaaact ggagtgagtg ttagcagaac ccgagcttag 120

aagttctcta ctttcgnttt cttaanngcc ctttgcctgn ntggaggann agttctccag 180

ncnnttcacg ctcannactc acagnncttc nntccaagnn ncatcanncc ctgngngnag 240

nnntttccnt gnnagnggnt nntttcntac nantnaanan ntgnagngng nnctgnncac 300

tantntgncn cntgntnntc tngnctnnnt cgcaangnnt attnncaant nanccngncn 360

tncacgntna ntntntntan annactagna nangcntngc antntncntn anangncact 420

nttcggcgct ctctcgngcg cactacnaca ctnangagna nanacgnnca ntnantngnc 480

nangncncna nancacnana g 501

<210> 15
<211> 569
<212> DNA
<213> Homo sapien

<400> 15
acagaacatg atcaagggtg ttacactggg cttccgttac aagatgaggt ctgtgtatgc 60

tcacttcccc atcaacgtat gttatccagg agaatgggtc tcttgtagaa atccgaaata 120

tcttgggtga aaaatatatc cgcagggttc ggatgagacc aggtgttgct tgttcagtat 180

ctcaagccca gaaagatgaa ttaatccttg aaggaaatga cattgagcta gtttcaaatt 240

cagcgtgctt tggatgtcag cagatgccac aatcagttaa gaacaaggat atcaggaaat 300

ttttggatgg tatctatgtc tctgaaaaag gaactgttca gcaggctgat gaataagatc 360

taagagttac ctggctacag aaagaagatg ccagatgaca cttaagacct acttgtgata 420

tttaaatgat gcaataaaag acccattgat ttggaccttc ttcttaaaaa aaaaaaaaca 480

aaaaaaaaaa aagccggggg aaaacagggg ccaaggggt cccgggtgga cattgtttcc 540

ggcccaattt cccacatttt ggacaaaat 569

<210> 16
<211> 971
<212> DNA
<213> Homo sapien

<400> 16
atgaagacta ttctcagcaa tcagactgtc gacattccag aaaatggtat gagacttgat 60

gtcttttact tacatcttta ctgcacgttc caagcgttgt gtggcctgac gagtgtgttc 120

tctcttctag tcgacattac tctgaaggga cgcacagtta tcgtgaaggg ccccagagga 180

accctgcgga gggacttcaa tcacatcaat gtagaactca gccttcttgg aaagaaaaaa 240

aagaggctcc gggttgacaa atggtggggt aacagaaagg aactggctac cgttcggact 300

atttgtagtc atgtacagaa catgatcaag ggtgttacac tgggcttccg ttacaagatg 360

aggtctgtgt atgctcactt ccccatcaac gttgttatcc aggagaatgg gtctcttgtt 420

gaaatccgaa atttcttggg tgaaaaatat atccgcaggg ttcggatgag accaggtgtt 480

gcttgttcag tatctcaagc ccagaaagat gaattaatcc ttgaaggaaa tgacattgag 540

cttgtttcaa attcagcggc tttgattcag caagccacaa cagttaaaaa caaggatatc 600

aggaaatttt tggatggtat ctatgtctct gaaaaaggaa ctgttcagca ggctgatgaa 660

taagatctaa gagttacctg gctacagaaa gaagatgcca gatgacactt aagacctact 720

tgtgatattt aaatgatgca ataaaagacc tattgatttg gaccttcttc ttaaaaaaag 780

aaaaaaaaga caaagaacaa catagagcaa aaacgagcaa gcaaaaaaca gaagaacaca 840

gccccgggcg attttattgt tgggcgggcg gcgcgaaacc agggcctcag tcaacggcca 900

ggttgccata ggggtgtccc gccccctttt ttttccccga gtgcgaacac ccggcgcccc 960

aatgagggac a 971

<210> 17
<211> 422
<212> DNA
<213> Homo sapien

<400> 17
acaactccaa aaggagacat tggagaagaa ccaagctggg tctataagga attgcacatg 60

agatggcaca catatttatg ctgtctgaag gtcacgatca tgttaccata tcaagctgaa 120

aatgtcacca ctatctggag atttcgacgt gttttcctct ctgaatctgt tatgaacacg 180

ttggttggct ggattcagta ataaatatgt aaggcctttc tttttaaaaa aaaacaacaa 240

aaaaaaaaaa aaaaaaaccc ctggggcgta ccccggggca aaagggtggt ccacggggtg 300

agacttggtt ccccggcgca aaatcccccc acactactaa gaacaagagg gccacggagg 360

agcagcacgc acagatcaca gcagaccgac acagatagca acacagagac acacacgcat 420

ag 422

<210> 18
<211> 584
<212> DNA
<213> Homo sapien

<400> 18
aagaattcac tagtaatcgc catcgtggtg tgttcttgac tccgctgctc gccatgtctt 60

ctcacaagac tttcaggatt aagcgattcc tggccaagaa acaaaagcaa aatcgtccca 120

ttccccagtg gattcggatg aaaactggaa ataaaatcag gtacaactcc aaaaggagac 180

attggagaag aaccaagctg ggtctataag gaattgcaca tgagatggca cacatattta 240

tgctgtctga aggtcacgat catgttacca tatcaagctg aaaatgtcac cactatctgg 300

agatttcgac gtgttttcct ctctgaatct gttatgaaca cgttggttgg ctggattcag 360

taataaatat gtaaggcctt tctttttaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaac 420

ccctggggcg taccccgggg caaaagggtg gtccacgggg tgagacttgg ttccccggcg 480

caaaatcccc ccacactact aagaacaaga gggccacgga ggagcagcac gcacagatca 540

cagcagaccg acacagatag caacacagag acacacacgc atag 584

<210> 19
<211> 747
<212> DNA
<213> Homo sapien

<400> 19
acaatattga acatttttct atatcctttg atatctgcaa gcctgatttt cagtagctgg 60

aaatggaaag gccaaattta ttatctaatt ttatacatta ggacatgtgt ataatgtcca 120

attttatact gttataagtc acactatgat gaacattttt gtacataact aaccatattt 180

cagttcattt ctttaggtta ttatatatcc acagatatga cattcaattc tataaaaatt 240

atgtacattt taatttattt tatttttgta catgggaagc tcctatctta actcattaaa 300

ttcaataaat tttgtatttc tacaacagaa agccaacaaa gggagttgtt agtacatatt 360

tccaggaatg aagttgtctg gatgcagcta atgcctccat agaactgaca gtgctgaatt 420

tacgaaatgg aaagagttct ggaaaagcaa gaaaaaaagt cttgtttgaa accccacgtc 480

tactgtaggc acagaaggga atggaggcat ctgagcattt tattttccat ctctacagca 540

cctcagaaca cctacatttt atttttttc ttctcagaaa tgtcttaata agaggactgc 600

agtgtactca agtttcccaa tgacagggta gggatgccaa ccttctcttt cattggcagc 660

tcatagtatc caagtttctc aaaaccctaa gccatcttat ttgttctttg gaactttgtg 720

gcctaccaca gtgcaatctc atcggtg 747

<210> 20
<211> 766
<212> DNA
<213> Homo sapien

<400> 20
acaatattga acatttttct atatcctttg atatctgcaa gcctgatttt cagtagctgg 60

aaatggaaag gccaaattta ttatctaatt ttatacatta ggacatgtgt ataatgtcca 120

attttatact gttataagtc acactatgat gaacattttt gtacataact aaccatattt 180

cagttcattt ctttaggtta ttatatatcc acagatatga cattcaattc tataaaaatt 240

atgtacattt taatttattt tatttttgta catgggaagc tcctatctta actcattaaa 300

ttcaataaat tttgtatttc tacaacagaa agccaacaaa gggagttgtt agtacatatt 360

tccaggaatg aagttgtctg gatgcagcta atgcctccat agaactgaca gtgctgaatt 420

tacgaaatgg aaagagttct ggaaaagcaa gaaaaaaagt cttgtttgaa accccacgtc 480

tactgtaggc acagaaggga atggaggcat ctgagcattt tattttccat ctctacagca 540

cctcagaaca cctacatttt atttttttc ttctcagaaa tgtcttaata agaggactgc 600

agtgtactca agtttcccaa tgacagggta gggatgccaa ccttctcttt cattggcagc 660

tcatagtatc caagtttctc aaaaccctaa gccatcttat ttgttctttg gaactttgtg 720

gcctaccaca gtgcaattct cattcggtgt ttaataactc gagccg 766

<210> 21
<211> 647
<212> DNA
<213> Homo sapien

<400> 21
tgaacatcat catgaataca tgaatcggct gtgatgtgtg aactgctaag ggccaaatga 60

acgtttgcag agcagtgggc acaatgttta caatgtatgt gtatgtcact ttcggtacct 120

gtgaatgcat ggggacgtgc tgaacccgaa aaaaagtgcc tttccataag gactgcaata 180

gagagggcaa tttaccctgg tggtacacgg aacctagatt cactcctgcc atgccttgcc 240

aatagtaagc tgcagggtgg aacaagaaat cacttgctct gggggaagg gagggggaa 300

tgggtgtgtc agctgggtag atacaaaccc tgaaaagaga atccatgtgc tgctggcagg 360

caacattttt taaagctctt tcagaaaccc tcatatttgg ggtttctttt caggaaacat 420

tcctgtggag ggaaaacgaa tatgaagata attttcagct aattatctgg gtgacccaga 480

atcgtgtata tggctatagg atagacttct taataatggc aagtgacgtg gccctgggga 540

aaggtgcttt atgtaccgtg tgtgcgtgta tgtgtgtgta tctatacaag tttgtcagct 600

ttggcatgac tgtttgtttg tctcgaaaac caataaactc aaagttt 647

<210> 22
<211> 698
<212> DNA
<213> Homo sapien

<400> 22
actagcaccg ggcaagcaga caacataatt tatttccaga aaacaacaga atgaacatca 60

tcatgaatac atgaatcggc tgtgatgtgt gaactgctaa gggccaaatg aacgtttgca 120

gagcagtggg cacaatgttt acaatgtatg tgtatgtcac tttcggtacc tgtgaatgca 180

tggggacgtg ctgaacccga aaaaaagtgc ctttccataa ggactgcaat agagagggca 240

atttaccctg gtggtacacg gaacctagat tcactcctgc catgccttgc caatagtaag 300

ctgcagggtg gaacaagaaa tcacttgctc tgggggggaag ggagggggga atgggtgtgt 360

cagctgggta gatacaaacc ctgaaaagag aatccatgtg ctgctggcag gcaacatttt 420

ttaaagctct ttcagaaacc ctcatatttg gggtttcttt tcaggaaaca ttcctgtgga 480

gggaaaacga atatgaagat aattttcagc taattatctg ggtgacccag aatcgtgtat 540

atggctatag gatagacttc ttaataatgg caagtgacgt ggccctgggg aaaggtgctt 600

tatgtaccgt gtgtgcgtgt atgtgtgtgt atctatacaa gtttgtcagc tttggcatga 660

ctgtttgttt gtctcgaaaa ccaataaact caaagttt 698

<210> 23
<211> 739
<212> DNA
<213> Homo sapien

<400> 23
taaacttaag gctaatgttt agaagctttt gctaatgaga ggaccatttg ctaaatcggt 60

ataagtgcta cacatttggg tatctccatc ccaacatacc tcttattgcc attccccaaa 120

gcagacaccg tctcctccct ccctcaagga cctctgagct tgcactccaa ttcctctccc 180

acactcacct ttctcctttc tgttcctctt gggatccagg tttatttgag gagataggaa 240

aagctcctga tccagcaggt tttattctta aatttgtaac aaagtaaatc acagaacctc 300

cacccagcat ccaggcctct ggttctctcc ctccttccca ggtataggcc ggctttcaga 360

aaccctgcac cacatagacc ctgggcctga attgctgtga gtaataatga ctctgctcgt 420

aatttgtgtc cttctgcttg gaactgtttc ctttttagtt tggtcaccct cccagagctg 480

gtttcaatgg gggcataccc attatgggat gcagggcatc ctgcatcctg aggaattttt 540

tttcctccaa aaatgaaacc ttgaaatgag gacattgtcc tgtccacgga ctgcacaaca 600

acactgagcc tcaaggactc atactggcat ttttcttctt ttgcagagtg tgggcaccct 660

ggcttcaagc tcacgagaaa ccaggtcggg atttaaacaa tgttgggtta aagcaaagtt 720

tcataaagac agaatcaag 739

<210> 24
<211> 900
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (75)..(75)
<223> n=a, c, g, or t

<400> 24
agcgacattc ggcacgagta cgtaatatac tccagtttgc aaatgaagga atcttcctgc 60

ggaacgtatg tgaangcata ttggtgctct gggcttttgc ataatttcaa atgtcctttt 120

tttttaaact taaggctaat gtgtagaagc ttttgctaat gagaggacca tttgctaaat 180

cggtataagt gctacacatt tgggtatctc catcccaaca tacctcttat tgccattccc 240

caaagcagac accttctcct ccctccctca aggacctctg agcttgcact ccaattcctc 300

tcccacactc acctttctcc tttctgttcc tcttgggatc caggtttatt tgaggagata 360

ggaaaagctc ctgatccagc aggttttatt cttaaatttg taacaaagta aatcacagaa 420

cctccaccca gcatccaggc ctctggttct ctccctcctt cccaggtata ggccggcttt 480

cagaaaccct gcaccacata gaccctgggc ctgaattgct gtgagtaata atgactctgc 540

tcgtaatttg tgtccttctg cttggaactg tttccttttt agtttggtca ccctcccaga 600

gctggtttca atgggggcat acccattatg ggatgcaggg catcctgcat cctgaggaat 660

tttttttcct ccaaaaatga aaccttgaaa tgaggacatt gtcctgtcca cggactgcac 720

aacaacactg agcctcaagg actcatactg gcatttttct tcttttgcag agtgtgggca 780

ccctggcttc aagctcacga gaaaccaggt cgggatttaa acaatgttgg gttaaagcaa 840

agtttcataa agacagaatc aagaaaaaaa aaaaaaaaaa atatactggc cgcaaggaat 900

<210> 25
<211> 299
<212> DNA
<213> Homo sapien

<400> 25

ggcagcgcgg aggccgcacg atgcctggag ttactgtaaa agacgtgaac cagcaggagt 60

tcgtcagagc tctggcagcc ttcctcaaaa agtccgggga agctgaaagt ccccgaatgg 120

gtgggatacc gttcaagctg gccaaagcac aaaggagctt gctccctacg atgagaactg 180

gttctacacg cggagctgct ttccaacagc ggcgggccac ctgttacctt ccgggggtgg 240

gcgctggggg ttgggcttcc attgaaccca aggattctat tgggggggaa cgttcagaa 299

<210> 26
<211> 1346
<212> DNA
<213> Homo sapien

<400> 26
tttttttttt ttgtgagcca gtgggaaaac caaggaggct aaaccataga gcctggagat 60

gtgaaggaag tacaggtggg taagaaaggg agagccagat cacaagcacc ttgaaaccag 120

acactggttt ggggtcttca gcagtcctct gtcgaaatac atatattcag gggctgggtg 180

tggtggctca cacctgtaat cccagccctt tgggaggcag aggcaggcag attacttgag 240

gtcaagagtt caagacaagc ctggccaaca tggtaaaacc ccgtctctac caaaaatata 300

aaaaactagc cgggcgtggt ggcaggcacc tgattgtaat cccagctact cgggaggctg 360

aggcaggaga atcatttgaa cccagaaggc ggagattgca gtgagctgag atggcgccac 420

tgccactccc agcctgggcg acagagcaag agactcaaaa aagagaccca gaccaggatt 480

acgaatgagg caatttatta acccagcatg gtttgttcta atgcttcttg ttggcagctg 540

ccacctgtcc ggcgattctg tccagatctc tttgtccctg aggtgtcagt ttgcggccgc 600

catcttggtc cttttccacc attttcagcc cctccagggc ttggaggacc cggcgggcca 660

cactcttgga gcctcggctg aagtggctgg gcatgacgcc gtttctctga cgtcccccat 720

agatcttggt catggagcca accccagcgc caccccggag gtacaggtgc cgcgctgtgg 780

aagcagctcg cgtgtagaac cagttctcat cgtagggagc aagctctttg tgcttggcca 840

gcttgacggt atccacccat tcggggactt tcagcttccc ggactttttg aggaaggctg 900

ccagagctct gacgaactcc tgctggttca cgtcttttac agtaactcca ggcatcgtgc 960

ggcctccgcg ctgccagcca ggggaaaggg aacgacgggg tttcccgggc gcacaagtcg 1020

ggcgtagggt ctcgcgagag ttccgaaagc tcgcgagagc gagggtagac gctgaggctc 1080

cgcctctctc agggcgaaag ttcgtccccg cctagagggg agggtgtcta gtgaggggtg 1140

gagaggtaaa ggggagggcc aaggggtcgc gcgtggaggc ctgggtttcc tcccgcgttt 1200

ccttctcccg gagtgtaata gagagaggat agagagctcc tgttcggagc tgggggaact 1260

tggcttcgtt tgcgtcgttc gtggctggaa ggaacagtgg tggagaatac tatgatggcg 1320

aaagtacggg gcaggatggg tgggcc 1346

<210> 27
<211> 136
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (75)..(75)
<223> n=a, c, g, or t

<400> 27
gtcgccgctg cgaagggagc cgccgccatg tctgcgcatc tgcaatggat ggtcgtgcgg 60

aactgctcca gtttnctgat caagaggaat agccccctgc ccccagagca ataaagtcag 120

ctggctttct cacctg 136

<210> 28
<211> 426
<212> DNA
<213> Homo sapien

<400> 28
gctcgaggcc atttcctctc tccagaggac ctttcctgcc taggactcat cattgtcccc 60

tccctggcat tttttacacc tggagcagcc agaggacgca tgcatggctc ttcggaagcc 120

ttctcctgcc acggcatgca cccacacatg cgagcctccc gggtactgtc atcctgaatt 180

ctgagaccat ccagcacttc ctttagtttt gccctggtgc tgttgacttt tgtttactga 240

agagtgtgct ggaggcagga caagggacat ggaaggctgc aatttaagag tctaaaaggt 300

tttagaatcc tgaaggaggt ttaacaagct gaattgaaga ataatacctt tctcaactgg 360

agagaattta catgattgca ttattgttaa aattaacatc tcatctatta aaagcatttg 420

tagatt 426

<210> 29
<211> 264
<212> DNA
<213> Homo sapien

<400> 29
cgggaaccct gagacctctc cagcgaagct gaagtgctgt gttacgggag agagtgactg 60

gaaagtaaca aagctgaatc tttctccctg gagtaaggcc gaagactgga ttactacacg 120

cctagacgtg acactacacc catagatctc atgcatcatt aatgccatat gacattgcca 180

ttttctttct cagttcacgg acaaaagtgg tgggttttca ttgcttcact gattgtcaat 240

gcattaataa agaagatgtg tggt 264

<210> 30
<211> 265
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (164)..(164)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (168)..(168)
<223> n=a, c, g, or t

<400> 30
cggccttcct gtaagaaaga tccacggccg ggcccgggcg gccccgcttc ccagagactc 60

atccagccgg aggagatgtg gctctaccgg aacccctacg tggaggcgga gtatttcccc 120

accaagccga tgtttgtgcg tggagaaaga tcgtctttcc tccntccnca tgacccggct 180

tcccgcgggc acctgtgcgt tttccacccc gagacggcct ttgttattgc atttctctct 240

ccactgtctc tgatcttcct ggcca 265

<210> 31
<211> 741
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (718)..(718)
<223> n=a, c, g, or t

<400> 31
ggcaaccaca ggttccaaga tggtttgcgg gggcttcgcg tgttccagtc tccgagtggt 60

cggcgtggtc attgcagtgg gcatcttctt gttcctgatt gctttagtgg gtctgattgg 120

agctgtaaaa catcatcagg tgttgctatt cttttatatg attattctgt tacttgtatt 180

tattgttcag ttttctgtat cttgcgcttg tttagccctg aaccaggagc aacagggtca 240

gcttctggag gttggttgga acaatacggc aagtgctcga aatgacatcc agagaaatct 300

aaactgctgt gggttccgca gtgttaaccc aaatgacacc tgtctggcta gctgtgttaa 360

aagtgaccac tcgtgctcgc catgtgctcc aatcatagga gaatatgctg gagaggtttt 420

gagatttgtt ggtggcattg gcctgttctt cagttttaca gagatcctgg gtgtttggct 480

gacctacaga tacaggaacc agaaagaccc ccgcgcgaat cctagtgcat tcctttgatg 540

agaaaacaag gaagatttcc tttcgtatta tgatcttgtt cactttctgt aattttctgt 600

```
taagctccat ttgccagttt aaggaaggaa acactatctg gaaaagtacc ttattgatag    660

tggaattata tatttaccta gtttctctac agttttcttc cgtgcgaaaa atattganac    720

tgggcctgaa ccggggcacg g                                              741


<210>  32
<211>  1844
<212>  DNA
<213>  Homo sapien

<400>  32
aaggatcctt aattaaatta atcccccccc cccgctcctt gccagcgtgg atctcctccg     60

agccccgccc tccctcctca cctgctcctg gggaaactac accaaggccg ccgctctggc    120

ctggggctcc ctcccacacg gccttggccc tctccccctc gccccgggac cgctccgccc    180

ctcccggatc ccggtcggcg gagcgcattt atttgcatat ttctaccttt gttccccgcc    240

tgggccaggc cccaaaggca aggacaaagc agctgtcagg gaacctccgc cggagtcgaa    300

tttacgtgca gctgccggca accacaggtt ccaagatggt ttgcgggggc ttcgcgtgtt    360

ccaagaactg cctgtgcgcc ctcaacctgc tttacacctt ggttagtctg ctgctaattg    420

gaattgctgc gtggggcatt ggcttcgggc tgatttccag tctccgagtg gtcggcgtgg    480

tcattgcagt gggcatcttc ttgttcctga ttgctttagt gggtctgatt ggagctgtaa    540

aacatcatca ggtgttgcta ttcttttata tgattattct gttacttgta tttattgttc    600

agttttctgt atcttgcgct tgtttagccc tgaaccagga gcaacagggt cagcttctgg    660

aggttggttg gaacaatacg gcaagtgctc gaaatgacat ccagagaaat ctaaactgct    720

gtgggttccg aagtgttaac ccaaatgaca cctgtctggc tagctgtgtt aaaagtgacc    780

actcgtgctc gccatgtgct ccaatcatag gagaatatgc tggagaggtt ttgagatttg    840

ttggtggcat tggcctgttc ttcagtttta cagagatcct gggtgtttgg ctgacctaca    900

gatacaggaa ccagaaagac ccccgcgcaa atcctagtgc attcctttga tgagaaaaca    960

aggaagattt cctttcgtat tatgatcttg ttcactttct gtaattttct gttaagctcc   1020

atttgccagt ttaaggaagg aaacactatc tggaaaagta ccttattgat agtggaatta   1080

tatattttta ctctatgttt ctctacatgt ttttttcttt ccgttgctga aaaatatttg   1140

aaacttgtgg tctctgaagc tcggtggcac ctggaattta ctgtattcat tgtcgggcac   1200

tgtccactgt ggcctttctt agcattttta cctgcagaaa aactttgtat ggtaccactg   1260

tgttggttat atggtgaatc tgaacgtaca tctcactggt ataattatat gtagcactgt   1320

gctgtgtaga tagttcctac tggaaaaaga gtggaaattt attaaaatca gaaagtatga   1380

gatcctgtta tgttaaggga aatccaaatt cccaattttt tttggtcttt ttaggaaaga   1440
```

tgtgttgtgg taaaaagtgt tagtataaaa atgataattt acttgtagtc ttttatgatt 1500

acaccaatgt attctagaaa tagttatgtc ttaggaaatt gtggtttaat ttttgacttt 1560

ttacaggtaa gtgccaagga gaagtggttc ctgaaatgtt ctaatgttta ttaacatttt 1620

aaccttcagc tccatcagaa tggaccgagt tgagtaatca ggaggataac tatatgatct 1680

gaatggtata ctaattggag ctaaagacgc ttttcaccag ttgtttattg gttggccgtg 1740

caaaagattt gttttcaaat gggaaacggg cgaattcgtt ggacgctgtg cagtttgttg 1800

tccctgagaa gatggggggt ttaaaagagg caaaaaaaaa aggg 1844

<210> 33
<211> 242
<212> DNA
<213> Homo sapien

<400> 33
gctctcactg cctgtgagag ccccatcgtg gtggtgctga gtggcaggag gtctcctctg 60

ttcctcacaa atttccggga agccccaatc agagctggtg aaatagttgt ttttaaagtt 120

gaaggacgag acattccaat agttcacaga gtaatcaaag ttcatgaaaa agataatgga 180

gacatcaaat ttctgactaa aggagataat aatgaagttg atgatagagg cttgtacaaa 240

ga 242

<210> 34
<211> 966
<212> DNA
<213> Homo sapien

<400> 34
aaggatcctt aattaaatta atcccccccc ccccggcagc cgtctgtgcc acccagagcc 60

ggcgggccgc taggtccccg gagaccctgc tatggtgcgt gcgggcgccg tgggggctca 120

tctccccgcg tccggcttgg atatcttcgg ggacctgaag aagatgaaca agcgccagct 180

ctattaccag gttttaaact tcgccatgat cgtgtcttct gcactcatga tatggaaagg 240

cttgatcgtg ctcacaggca gtgagagccc catcgtggtg gtgctgagtg gcagtatgga 300

gccggccttt cacagaggag acctcctgtt cctcacaaat ttccgggaag acccaatcag 360

agctggtgaa atagttgttt ttaaagttga aggacgagac attccaatag ttcacagagt 420

aatcaaagtt catgaaaaag ataatggaga catcaaattt ctgactaaag gagataataa 480

tgaagttgat gatagaggct tgtacaaaga aggccagaac tggctggaaa agaaggacgt 540

ggtgggaaga gcaagagggt ttttaccata tgttggtatg gtcaccataa taatgaatga 600

ctatccaaaa ttcaagtatg ctcttttggc tgtaatgggt gcatatgtgt tactaaaacg 660

tgaatcctaa aatgagaagc agttcctggg accagattga aatgaattct gttgaaaaag 720

agaaaaacta atatatttga gatgttccat tttctgtata aaagggaaca gtgtggaaat 780

tgtccgcggt cttgggccaa gtaatagatt tgccgcgggg aaggaaatgg gagtttgtta 840

taaagatggt gcggcagctt ggaggctgtg ctgttccctt cgagttgggg ccgaataatc 900

gaccatgtgt gcccttcctc gcgtccttct agctatgcgg gcgctatgaa ccgggcggtt 960

gggttt 966

<210> 35
<211> 717
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (685)..(686)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (688)..(688)
<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (697)..(697)
<223> n=a, c, g, or t

<400> 35
catgacaccg ggcacccagt ctcctttctt cctgctgctg ctcctcacag tgcttacagt 60

tgttacaggt tctggtcatg caagctctac cccaggtgga gaaaaggaga cttcggctac 120

ccagagaagt tcagtgccca gctctactga gaagaatgct tttaattcct ctctggaaga 180

tcccagcacc gactactacc aagagctgca gagagacatt tctgaaatgg ctgtctgtca 240

gtgccgccga aagaactacg ggcagctgga catctttcca gccccgggat acctaccatc 300

ctatgagcga gtaccccacc taccacaccc atgggcgcta tgtgcccccct agcagtaccg 360

atcgtagccc ctatgagaag gtttctgcag gtaatggtgg cagcagcctc tcttacacaa 420

acccagcagt ggcagccact tctgccaact tgtaggggca cgtcgcccgc tgagctgagt 480

ggccagccag tgccattcca ctccactcag gttcttcagg gccagagccc ctgcacctgt 540

ttgggctggt gagctgggag ttcaggtggg ctgctcacag cctccttcag aggccccacc 600

aatttctcgg acacttctca gtgtgtggaa gctcatgtgg gcccctgagg gctcatgcct 660

gggaagtgtt gtggtgggtg ctacnnanga ggactgnccc agagagccct gagatag 717

<210> 36
<211> 774
<212> DNA
<213> Homo sapien

<400> 36
catgacaccg ggcacccagt ctcctttctt cctgctgctg ctcctcacag tgcttacagt 60

tgttacaggt tctggtcatg caagctctac cccaggtgga gaaaaggaga cttcggctac 120

ccagagaagt tcagtgccca gctctactga gaagaatgct tttaattcct ctctggaaga 180

tcccagcacc gactactacc aagagctgca gagagacatt tctgaaatgg ctgtctgtca 240

gtgccgccga aagaactacg ggcagctgga catctttcca gccccgggat acctaccatc 300

ctatgagcga gtaccccacc taccacaccc atgggcgcta tgtgccccct agcagtaccg 360

atcgtagccc ctatgagaag gtttctgcag gtaatggtgg cagcagcctc tcttacacaa 420

acccagcagt ggcagccact tctgccaact tgtaggggca cgtcgcccgc tgagctgagt 480

ggccagccag tgccattcca ctccactcag gttcttcagg gccagagccc ctgcaccctg 540

tttgggctgg tgagctggga gttcaggtgg gctgctcaca gcctccttca gaggccccac 600

caatttctcg gacacttctc agtgtgtgga agctcatgtg ggcccctgag ggctcatgcc 660

tgggaagtgt tgtggtgggt gctcccagga ggactggccc agagagccct gagatagcgg 720

ggatcctgaa ctggactgaa taaaacgtgg tctcccctg cgccaaaaaa aaaa 774

<210> 37
<211> 4144
<212> DNA
<213> Homo sapien

<400> 37
ccgctccacc tctcaagaat tccctggctg cttgaatctg ttctgccccc tccccaccca 60

tttcaccacc accatgacac cgggcaccca gtctcctttc ttcctgctgc tgctcctcac 120

agtgcttaca gttgttacag gttctggtca tgcaagctct accccaggtg gagaaaagga 180

gacttcggct acccagagaa gttcagtgcc cagctctact gagaagaatg ctgtgagtat 240

gaccagcagc gtactctcca gccacagccc cggttcaggc tcctccacca ctcagggaca 300

ggatgtcact ctggccccgg ccacggaacc agcttcaggt tcagctgcca cctggggaca 360

ggatgtcacc tcggtcccag tcaccaggcc agccctgggc tccaccaccc cgccagccca 420

cgatgtcacc tcagccccgg acaacaagcc agccccgggc tccaccgccc ccccagccca 480

cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgccc ccccagccca 540

cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgcca gcccacggtg 600

tcacctcggc cccggacacc aggccgsccc cgggctccac cgcsccсsca gcccacggtg 660

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccayggtg 720

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 780

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 840

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 900

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 960

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1020

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1080

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1140

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1200

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1260

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1320

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1380

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1440

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1500

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1560

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1620

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1680

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1740

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1800

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1860

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1920

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 1980

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2040

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2100

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2160

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2220

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2280

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2340

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2400

tcacctcggc cccggacacc aggccggccc cgggctccac cgcccccca gcccacggtg 2460

tcacctcggc cccggacacc aggccggccc cgggctccac cgccccccca gcccacggtg 2520

tcacctcggc cccggacacc aggccggccc cgggctccac cgccccccca gcccacggtg 2580

tcacctcggc cccggacacc aggccggccc cgggctccac cgccccccca gcccacggtg 2640

tcacctcggc cccggacacc aggccggccc cgggctccac cgccccccca gcccacggtg 2700

tcacctcggc cccggacacc aggccggccc cgggctccac cgccccccca gtccacggtg 2760

tcacctcggc cccggactcc aggtcgggct cgggcttcct accgccgccc gcagcccacg 2820

gtgtcacctc ggccccggac accaggccgg ccccgggctc caccgccccc ccagcccatg 2880

gtgtcacctc ggccccggac aacaggcccg ccttggcgct ccaccgcccc tccagtccac 2940

aatgtcacct cggcctcagg ctctgcatca ggctcagctt ctactctggt gcacaacggc 3000

acctctgcca gggctaccac aaccccagcc agcaagagca ctccattctc aattcccagc 3060

caccactctg atactcctac caccccttgcc agccatagca ccaagactga tgccagtagc 3120

actcaccata gcacggtacc tcctctcacc tcctccaatc acagcacttc tccccagttg 3180

tctactgggg tctctttctt tttcctgtct tttcacattt caaacctcca gtttaattcc 3240

tctctggaag atcccagcac cgactactac caagagctgc agagagacat ttctgaaatg 3300

tttttgcaga tttataaaca agggggtttt ctgggcctct ccaatattaa gttcaggcca 3360

ggatctgtgg tggtacaatt gactctggcc ttccgagaag gtaccatcaa tgtccacgac 3420

gtggagacac agttcaatca gtataaaacg gaagcagcct ctcgatataa cctgacgatc 3480

tcagacgtca gcgtgagtga tgtgccattt cctttctctg cccagtctgg ggctggggtg 3540

ccaggctggg gcatcgcgct gctggtgctg gtctgtgttc tggttgcgct ggccattgtc 3600

tatctcattg ccttggctgt ctgtcagtgc cgccgaaaga actacgggca gctggacatc 3660

tttccagccc gggataccta ccatcctatg agcgagtacc ccacctacca cacccatggg 3720

cgctatgtgc cccctagcag taccgatcgt agcccctatg agaaggtttc tgcaggtaat 3780

ggtggcagca gcctctctta cacaaaccca gcagtggcag ccacttctgc caacttgtag 3840

gggcacgtcg cccgctgagc tgagtggcca gccagtgcca ttccactcca ctcaggttct 3900

tcagggccag agcccctgca ccctgtttgg gctggtgagc tgggagttca ggtgggctgc 3960

tcacagcctc cttcagaggc cccaccaatt tctcggacac ttctcagtgt gtggaagctc 4020

atgtgggccc ctgaggctca tgcctgggaa gtgttgtggg ggctcccagg aggactggcc 4080

cagagagccc tgagatagcg gggatcctga actggactga ataaaacgtg gtctcccact 4140

gcga 4144

<210> 38
<211> 2255

<212> DNA
<213> Homo sapien

<400> 38
ccgctccacc tctcaagaat tccctggctg cttgaatctg ttctgccccc tccccaccca 60
tttcaccacc accatgacac cgggcaccca gtctcctttc ttcctgctgc tgctcctcac 120
agtgcttaca gttgttacag gttctggtca tgcaagctct accccaggtg gagaaaagga 180
gacttcggct acccagagaa gttcagtgcc cagctctact gagaagaatg ctgtgagtat 240
gaccagcagc gtactctcca gccacagccc cggttcaggc tcctccacca ctcagggaca 300
ggatgtcact ctggccccgg ccacggaacc agcttcaggt tcagctgcca cctggggaca 360
ggatgtcacc tcggtcccag tcaccaggcc agccctgggc tccaccaccc cgccagccca 420
cgatgtcacc tcagccccgg acaacaagcc agccccgggc tccaccgccc ccccagccca 480
cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgccc ccccagccca 540
cggtgtcacc tcggccccgg acaccaggcc gscccgggc tccaccgcsc ccscagccca 600
cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgccc ccccagccca 660
yggtgtcacc tcggccccgg acaacaggcc cgccttggcg ctccaccgcc cctccagtcc 720
acaatgtcac ctcggcctca ggctctgcat caggctcagc ttctactctg gtgcacaacg 780
gcacctctgc cagggctacc acaaccccag ccagcaagag cactccattc tcaattccca 840
gccaccactc tgatactcct accacccttg ccagccatag caccaagact gatgccagta 900
gcactcacca tagcacggta cctcctctca cctcctccaa tcacagcact tctccccagt 960
tgtctactgg ggtctctttc tttttcctgt cttttcacat ttcaaacctc cagtttaatt 1020
cctctctgga agatcccagc accgactact accaagagct gcagagagac atttctgaaa 1080
tggtgagtat cggcctttcc ttccccatgc tcccctgaag cagccatcag aactgtccac 1140
accctttgca tcaagcctga gtcctttccc tctcacccca gtttttgcag atttataaac 1200
aagggggttt tctgggcctc tccaatatta agttcaggta cagttctggg tgtggaccca 1260
gtgtggtggt tggagggttg ggtggtggtc atgaccgtag gagggactgg tcgcacttaa 1320
ggttggggga agagtcgtga gccagagctg ggacccgtgg ctgaagtgcc catttccctg 1380
tgaccaggcc aggatctgtg gtggtacaat tgactctggc cttccgagaa ggtaccatca 1440
atgtccacga cgtggagaca cagttcaatc agtataaaac ggaagcagcc tctcgatata 1500
acctgacgat ctcagacgtc agcggtgagg ctacttccct ggctgcagcc cagcaccatg 1560
ccggggccct ctccttccag tgcctgggtc cccgctcttt ccttagtgct ggcagcggga 1620
ggggcgcctc ctctgggaga ctgccctgac cactgctttt ccttttagtg agtgatgtgc 1680
catttccttt ctctgcccag tctggggctg gggtgccagg ctggggcatc gcgctgctgg 1740

tgctggtctg tgttctggtt gcgctggcca ttgtctatct cattgccttg gtgagtgcag 1800
tccctggccc tgatcagagc cccccggtag aaggcactcc atggcctgcc ataacctcct 1860
atctccccag gctgtctgtc agtgccgccg aaagaactac gggcagctgg acatctttcc 1920
agcccgggat acctaccatc ctatgagcga gtaccccacc taccacaccc atgggcgcta 1980
tgtgccccct agcagtaccg atcgtagccc ctatgagaag gtgagattgg ccccacaggc 2040
caggggaagc agagggtttg gctgggcaag gattctgaag gggtacttg gaaaacccaa 2100
agagcttgga agaggtgaga agtggcgtga agtgagcagg ggagggcctg gaaaggatga 2160
ggggcagagg tcagaggagt tttgggggac aggcctggga ggagactatg gaagaaaggg 2220
gccctcaaaa gggagtggcc ccactgccag aattc 2255

<210> 39
<211> 1953
<212> DNA
<213> Homo sapien

<400> 39
ccgctccacc tctcaagaat tccctggctg cttgaatctg ttctgccccc tccccaccca 60
tttcaccacc accatgacac cgggcaccca gtctcctttc ttcctgctgc tgctcctcac 120
agtgcttaca gctaccacag cccctacacc cgcaacagtt gttacaggtt ctggtcatgc 180
aagctctacc ccaggtggag aaaaggagac ttcggctacc cagagaagtt cagtgcccag 240
ctctactgag aagaatgctg tgagtatgac cagcagcgta ctctccagcc acagccccgg 300
ttcaggctcc tccaccactc agggacagga tgtcactctg gccccggcca cggaaccagc 360
ttcaggttca gctgccacct ggggacagga tgtcacctcg gtcccagtca ccaggccagc 420
cctgggctcc accaccccgc cagcccacga tgtcacctca gccccggaca acaagccagc 480
cccgggctcc accgcccccc cagcccacgg tgtcacctcg gccccggaca ccaggccggc 540
cccgggctcc accgcccccc cagcccacgg tgtcacctcg gccccggaca ccaggccggc 600
cccgggctcc accgcgcccg cagcccacgg tgtcacctcg gccccggaca ccaggccggc 660
cccgggctcc accgcccccc cagcccatgg tgtcacctcg gccccggaca acaggcccgc 720
cttggcgctc caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag 780
gctcagcttc tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca 840
gcaagagcac tccattctca attcccagcc accactctga tactcctacc acccttgcca 900
gccatagcac caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct 960
cctccaatca cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt 1020
ttcacatttc aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc 1080

aagagctgca gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc 1140

tgggcctctc caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct 1200

tccgagaagg taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg 1260

aagcagcctc tcgatataac ctgacgatct cagacgtcag cgtgagtgat gtgccatttc 1320

ctttctctgc ccagtctggg gctggggtgc caggctgggg catcgcgctg ctggtgctgg 1380

tctgtgttct ggttgcgctg gccattgtct atctcattgc cttggctgtc tgtcagtgcc 1440

gccgaaagaa ctacgggcag ctggacatct ttccagcccg ggatacctac catcctatga 1500

gcgagtaccc cacctaccac acccatgggc gctatgtgcc ccctagcagt accgatcgta 1560

gcccctatga gaaggtttct gcaggtaatg gtggcagcag cctctcttac acaaacccag 1620

cagtggcagc cacttctgcc aacttgtagg ggcacgtcgc ccgctgagct gagtggccag 1680

ccagtgccat tccactccac tcaggttctt cagggccaga gcccctgcac cctgtttggg 1740

ctggtgagct gggagttcag gtgggctgct cacagcctcc ttcagaggcc ccaccaattt 1800

ctcggacact tctcagtgtg tggaagctca tgtgggcccc tgaggctcat gcctgggaag 1860

tgttgtgggg gctcccagga ggactggccc agagagccct gagatagcgg ggatcctgaa 1920

ctggactgaa taaaacgtgg tctcccactg cga 1953

<210> 40
<211> 1738
<212> DNA
<213> Homo sapien

<400> 40
ccgctccacc tctcaagaat tccctggctg cttgaatctg ttctgccccc tccccaccca 60

tttcaccacc accatgacac cgggcaccca gtctcctttc ttcctgctgc tgctcctcac 120

agtgcttaca gttgttacag gttctggtca tgcaagctct accccaggtg gagaaaagga 180

gacttcggct acccagagaa gttcagtgcc cagctctact gagaagaatg ctgtgagtat 240

gaccagcagc gtactctcca gccacagccc cggttcaggc tcctccacca ctcagggaca 300

ggatgtcact ctggccccgg ccacggaacc agcttcaggt tcagctgcca cctggggaca 360

ggatgtcacc tcggtcccag tcaccaggcc agccctgggc tccaccaccc cgccagccca 420

cgatgtcacc tcagccccgg acaacaagcc agccccgggc tccaccgccc ccccagccca 480

cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgccc ccccagccca 540

cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgcgc ccgcagccca 600

cggtgtcacc tcggccccgg acaccaggcc ggccccgggc tccaccgccc ccccagccca 660

tggtgtcacc tcggccccgg acaacaggcc cgccttggcg ctccaccgcc cctccagtcc 720

acaatgtcac ctcggcctca ggctctgcat caggctcagc ttctactctg gtgcacaacg 780

gcacctctgc cagggctacc acaaccccag ccagcaagag cactccattc tcaattccca 840

gccaccactc tgatactcct accacccttg ccagccatag caccaagact gatgccagta 900

gcactcacca tagcacggta cctcctctca cctcctccaa tcacagcact tctccccagt 960

tgtctactgg ggtctctttc tttttcctgt cttttcacat ttcaaacctc cagtttaatt 1020

cctctctgga agatcccagc accgactact accaagagct gcagagagac atttctgaaa 1080

tgtttttgca gatttataaa caagggggtt ttctgggcct ctccaatatt aagttcaggc 1140

caggatctgt ggtggtacaa ttgactctgg ccttccgaga aggtaccatc aatgtccacg 1200

acgtggagac acagttcaat cagtataaaa cggaagcagc ctctcgatat aacctgacga 1260

tctcagacgt cagcgtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg 1320

tgccaggctg gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg 1380

tctatctcat tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca 1440

tctttccagc ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg 1500

ggcgctatgt gccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta 1560

atggtggcag cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt 1620

aggggcacgt cgcccgctga gctgagtggc cagccagtgc cattccactc cactcagggc 1680

tctctgggcc agtcctcctg ggagccccca ccacaacact tcccaggcat ggaattcc 1738

<210> 41
<211> 328
<212> DNA
<213> Homo sapien

<400> 41
tcatctcgag cggcggcgca gtgtgaggcg gcccgggctc accgcgcccg cagcccacgg 60

tgtcacctcg gccccggaca ccaggccggc cccgggctcc accgcgcccg cagcccacgg 120

tgtcacctcg gccccggaca ccaggccggc cccgggctcc accgcccccc cagcccatgg 180

tgtcaccttc gtgccycgga cctcaggtcg gcgccttgct ctctttctgg tctatgtgtt 240

ccgtgtagta agatgtagtt cagacgcgtc tcgatacact acgcatagcg aagtatatcg 300

atggatcata cgctgtttcc gtgtgtga 328

<210> 42
<211> 1030
<212> DNA
<213> Homo sapien

<220>

<221> misc_feature
<222> (574)..(574)
<223> n=a, c, g, or t

<400> 42
ccgctccacc tctcaagaat tccctggctg cttgaatctg ttctgccccc tccccaccca 60

tttcaccacc accatgacac cgggcaccca gtctcctttc ttcctgctgc tgctcctcac 120

agtgcttaca gctaccacag cccctacacc cgcaacagtt gttacaggtt ctggtcatgc 180

aagctctacc ccaggtggag aaaaggagac ttcggctacc cagagaagtt cagtgcccag 240

ctctactgag aagaatgctg tgagtatgac cagcagcgta ctctccagcc acagccccgg 300

ttcaggctcc tccaccactc agggacagga tgtcactctg gccccggcca cggaaccagc 360

ttcaggttca gctgccacct ggggacagga tgtcacctcg gtcccagtca ccaggccagc 420

cctgggctcc accacccccgc cagcccacga tgtcacctca gccccggaca acaagccagc 480

cccgggctcc accgccccgc ggccgatctt gtggctcggg cttgggtacc gcgtgcgtgc 540

ccggtcttca gctgcttcta gtaggtgctc accntacgca gttactaact tacgactgag 600

cgctgtcgct ttgcactaga cgatcgtgaa ctgggaacac ctcatgtgct gtcatcacaa 660

tttattcgct ttgcggcgcg atccccctgt tcgcaagagg gtggaagagg ccactgtgtg 720

taccccgcga acttagatcg tcggcggtgc tagactagat caccccctttg cgcagagact 780

gagagtattg gggacccaga aaacagaagc tgggggttca ggagttttgc acgacaaaga 840

actacgatag cagaagactt gatggtactg gtgacccaag gagaaatctg gggatttaga 900

ggccacctga aagatacgaa gatacaaata cagtctgaga tgctggggac ccaggagaca 960

gaggtggaca gcttctaggg taccagagtc agaggctgag ggggacagaa cgctaaaata 1020

ttagggaccc 1030

<210> 43
<211> 1918
<212> DNA
<213> Homo sapien

<400> 43
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctgtgagta tgaccagcag 360

```
cgtactctcc agccacagcc ccggttcagg ctcctccacc actcagggac aggatgtcac     420
tctggccccg gccacggaac cagcttcagg ttcagctgcc acctggggac aggatgtcac     480
ctcggtccca gtcaccaggc cagccctggg ctccaccacc ccgccagccc acgatgtcac     540
ctcagccccg gacaacaagc cagccccggg ctccaccgcc cccccagccc acggtgtcac     600
ctcggccccg gacaccaggc cggccccggg ctccaccgcc cccccagccc atggtgtcac     660
ctcggccccg gacaacaggc ccgccttggg ctccaccgcc cctccagtcc acaatgtcac     720
ctcggcctca ggctctgcat caggctcagc ttctactctg gtgcacaacg gcacctctgc     780
cagggctacc acaaccccag ccagcaagag cactccattc tcaattccca gccaccactc     840
tgatactcct accacccttg ccagccatag caccaagact gatgccagta gcactcacca     900
tagcacggta cctcctctca cctcctccaa tcacagcact tctccccagt tgtctactgg     960
ggtctctttc tttttcctgt cttttcacat ttcaaacctc cagtttaatt cctctctgga    1020
agatcccagc accgactact accaagagct gcagagagac atttctgaaa tgtttttgca    1080
gatttataaa caagggggtt ttctgggcct ctccaatatt aagttcaggc caggatctgt    1140
ggtggtacaa ttgactctgg ccttccgaga aggtaccatc aatgtccacg acgtggagac    1200
acagttcaat cagtataaaa cggaagcagc ctctcgatat aacctgacga tctcagacgt    1260
cagcgtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg tgccaggctg    1320
gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg tctatctcat    1380
tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca tctttccagc    1440
ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg ggcgctatgt    1500
gcccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta atggtggcag    1560
cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt aggggcacgt    1620
cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt cttcagggcc    1680
agagcccctg caccctgttt gggctggtga gctgggagtt caggtgggct gctcacagcc    1740
tccttcagag gccccaccaa tttctcggac acttctcagt gtgtggaagc tcatgtgggc    1800
ccctgagggc tcatgcctgg gaagtgttgt ggtgggggct cccaggagga ctggcccaga    1860
gagccctgag atagcgggga tcctgaactg gactgaataa aacgtggtct cccactgc      1918
```

<210> 44
<211> 1755
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1682)..(1682)

<223> n=a, c, g, or t

<220>
<221> misc_feature
<222> (1733)..(1733)
<223> n=a, c, g, or t

<400> 44
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctgtgagta tgaccagcag 360

cgtactctcc agccacagcc ccggttcagg ctcctccacc actcagggac aggatgtcac 420

tctggccccg gccacggaac cagcttcagg ttcagctgcc acctggggac aggatgtcac 480

ctcggtccca gtcaccaggc cagccctggg ctccaccacc ccgccagccc acgatgtcac 540

ctcagccccg gacaacaagc cagccccggg ctccaccgcc cccccagccc acggtgtcac 600

ctcggccccg gacaccaggc cggccccggg ctccaccgcc cccccagccc atggtgtcac 660

ctcggccccg gacaacaggc ccgccttggg ctccaccgcc cctccagtcc acaatgtcac 720

ctcggcctca ggctctgcat caggctcagc ttctactctg gtgcacaacg gcacctctgc 780

cagggctacc acaaccccag ccagcaagag cactccattc tcaattccca gccaccactc 840

tgatactcct accacccttg ccagccatag caccaagact gatgccagta gcactcacca 900

tagcacggta cctcctctca cctcctccaa tcacagcact tctccccagt tgtctactgg 960

ggtctctttc tttttcctgt cttttcacat ttcaaacctc cagtttaatt cctctctgga 1020

agatcccagc accgactact accaagagct gcagagagac atttctgaaa tgtttttgca 1080

gatttataaa caagggggtt ttctgggcct ctccaatatt aagttcaggc caggatctgt 1140

ggtggtacaa ttgactctgg ccttccgaga aggtaccatc aatgtccacg acgtggagac 1200

acagttcaat cagtataaaa cggaagcagc ctctcgatat aacctgacga tctcagacgt 1260

cagcgtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg tgccaggctg 1320

gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg tctatctcat 1380

tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca tctttccagc 1440

ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg ggcgctatgt 1500

gccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta atggtggcag 1560

cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt aggggcacgt 1620

cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt cttcaggcag 1680

ancctgacct gttggctgta gctggagtca gtggtgtaag ctcttcaagg ggncagtcat 1740

cgatatgtaa cgttc 1755

<210> 45
<211> 1530
<212> DNA
<213> Homo sapien

<400> 45
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gcttttaatt cctctctgga 360

agatcccagc accgactact accaagagct gcagagagac atttctgaaa tgtttttgca 420

gatttataaa caagggggtt ttctgggcct ctccaatatt aagttcaggc caggatctgt 480

ggtggtacaa ttgactctgg ccttccgaga aggtaccatc aatgtccacg acgtggagac 540

acagttcaat cagtataaaa cggaagcagc ctctcgatat aacctgacga tctcagacgt 600

cagcgtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg tgccaggctg 660

gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg tctatctcat 720

tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca tctttccagc 780

ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg ggcgctatgt 840

gccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta atggtggcag 900

cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt aggggcacgt 960

cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt cttcagggcc 1020

agagcccctg caccctgttt gggctggtga gctgggagtt caggtgggct gctcacagcc 1080

tccttcagag gccccacgac tatttcagga agttcgaacc ccacctgtac tccctcgact 1140

ccaacagcga cgatgtggac tctctgacag acgaggagat cctgtccaag taccagctgg 1200

gcatgctgca cttcagcact cagtacgacc tgctgcacaa ccacctcacc gtgcgcgtga 1260

tcgaggccag ggacctgcca cctcccatct cccacgatgg ctcgcgccag gacatggcgc 1320

actccaaccc ctacgtcaag atctgtctcc tgccagacca gaagaactca aagcagaccg 1380

gggtcaaacg caagacccag aagcccgtgt ttgaggagcg ctacaccttc gagatcccct 1440

tcctggaggc ccagaggagg accctgctcc tgaccgtggt ggattttgat aagttctccc 1500

gccactgtgt cattgggaaa gtttctgtgg 1530

<210> 46
<211> 563
<212> DNA
<213> Homo sapien

<400> 46
ttttgctttt ttgcacccag aggcaaaatg ggtggagcac tatgcccagg ggagcccttc 60

ccgaggagtc ccaggggtga gcctctgtgc ccctaatcat ctcctaggaa tggagggtag 120

accgagaaag gctggcatag gggaggttt cccaggtaga agaagaagtg tcagcagacc 180

aggtttctgc aggtaatggt ggcagcagcc tctcttacac aaacccagca gtggcagcca 240

cttctgccaa cttgtagggg cacgtcgccc gctgagctga gtggccagcc agtgccattc 300

cactccactc aggttcttca gggccagagc ccctgcaccc tgtttgggct ggtgagctgg 360

gagttcaggt gggctgctca cagcctcctt cagaggcccc accaatttct cggacacttc 420

tcagtgtgtg gaagctcatg tgggcccctg agggctcatg cctgggaagt gttgtggtgg 480

gggctcccag gaggactggc ccagagagcc ctgagatagc ggggatcctg aactggactg 540

aataaaacgt ggtctcccac tgc 563

<210> 47
<211> 1945
<212> DNA
<213> Homo sapien

<400> 47
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cgtgagtgat gtgccatttc ctttctctgc 1320

ccagtctggg gctggggtgc caggctgggg catcgcgctg ctggtgctgg tctgtgttct 1380

ggttgcgctg gccattgtct atctcattgc cttggctgtc tgtcagtgcc gccgaaagaa 1440

ctacgggcag ctggacatct ttccagcccg ggatacctac catcctatga gcgagtaccc 1500

cacctaccac acccatgggc gctatgtgcc ccctagcagt accgatcgta gcccctatga 1560

gaaggtttct gcaggtaatg gtggcagcag cctctcttac acaaacccag cagtggcagc 1620

cacttctgcc aacttgtagg ggcacgtcgc ccgctgagct gagtggccag ccagtgccat 1680

tccactccac tcaggttctt cagggccaga gcccctgcac cctgtttggg ctggtgagct 1740

gggagttcag gtgggctgct cacagcctcc ttcagaggcc ccaccaattt ctcggacact 1800

tctcagtgtg tggaagctca tgtgggcccc tgagggctca tgcctgggaa gtgttgtggt 1860

gggggctccc aggaggactg gcccagagag ccctgagata gcggggatcc tgaactggac 1920

tgaataaaac gtggtctccc actgc 1945

<210> 48
<211> 1882
<212> DNA
<213> Homo sapien

<400> 48
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

aggtggagaa aaggagactt cggctaccca gagaagttca gtgcccagct ctactgagaa 300

gaatgctgtg agtatgacca gcagcgtact ctccagccac agccccggtt caggctcctc 360

caccactcag ggacaggatg tcactctggc cccggccacg gaaccagctt caggttcagc 420

tgccacctgg ggacaggatg tcacctcggt cccagtcacc aggccagccc tgggctccac 480

caccccgcca gcccacgatg tcacctcagc cccggacaac aagccagccc cgggctccac 540

cgcccccccca gcccacggtg tcacctcggc cccggacacc aggccggccc cgggctccac 600

cgcccccccca gcccatggtg tcacctcggc cccggacaac aggcccgcct tgggctccac 660

cgcccctcca gtccacaatg tcacctcggc ctcaggctct gcatcaggct cagcttctac 720

tctggtgcac aacggcacct ctgccagggc taccacaacc ccagccagca agagcactcc 780

attctcaatt cccagccacc actctgatac tcctaccacc cttgccagcc atagcaccaa 840

gactgatgcc agtagcactc accatagcac ggtacctcct ctcacctcct ccaatcacag 900

cacttctccc cagttgtcta ctggggtctc tttctttttc ctgtcttttc acatttcaaa 960

cctccagttt aattcctctc tggaagatcc cagcaccgac tactaccaag agctgcagag 1020

agacatttct gaaatgtttt tgcagattta taaacaaggg ggttttctgg gcctctccaa 1080

tattaagttc aggccaggat ctgtggtggt acaattgact ctggccttcc gagaaggtac 1140

catcaatgtc cacgacgtgg agacacagtt caatcagtat aaaacggaag cagcctctcg 1200

atataacctg acgatctcag acgtcagcgt gagtgatgtg ccatttcctt tctctgccca 1260

gtctggggct ggggtgccag gctggggcat cgcgctgctg gtgctggtct gtgttctggt 1320

tgcgctggcc attgtctatc tcattgcctt ggctgtctgt cagtgccgcc gaaagaacta 1380

cgggcagctg gacatctttc cagcccggga tacctaccat cctatgagcg agtaccccac 1440

ctaccacacc catgggcgct atgtgccccc tagcagtacc gatcgtagcc cctatgagaa 1500

ggtttctgca ggtaatggtg gcagcagcct ctcttacaca aacccagcag tggcagccac 1560

ttctgccaac ttgtaggggc acgtcgcccg ctgagctgag tggccagcca gtgccattcc 1620

actccactca ggttcttcag ggccagagcc cctgcaccct gtttgggctg gtgagctggg 1680

agttcaggtg ggctgctcac agcctccttc agaggcccca ccaatttctc ggacacttct 1740

cagtgtgtgg aagctcatgt gggcccctga gggctcatgc ctgggaagtg ttgtggtggg 1800

ggctcccagg aggactggcc cagagagccc tgagatagcg ggatcctga actggactga 1860

ataaaacgtg gtctcccact gc 1882

<210> 49
<211> 1930
<212> DNA
<213> Homo sapien

<400> 49
gtcgctctag aggacccctc ataggttcgc agggccatga gccaaggcct atgggcagag 60
agaaggaggc tgctgcaggg aaggaggcgg ccaacccagg ggttactgag gctgcccact 120
ccccagtcct cctggtatta tttctctggt ggccagagct tatattttct tcttgctctt 180
atttttcctt cataaagacc caaccctatg actttaactt cttacagcta ccacagcccc 240
taaacccgca acagttgtta caggttctgg tcatgcaagc tctaccccag gtggagaaaa 300
ggagacttcg gctacccaga gaagttcagt gcccagctct actgagaaga atgctgtgag 360
tatgaccagc agcgtactct ccagccacag ccccggttca ggctcctcca ccactcaggg 420
acaggatgtc actctggccc cggccacgga accagcttca ggttcagctg ccacctgggg 480
acaggatgtc acctcggtcc cagtcaccag gccagccctg ggctccacca ccccgccagc 540
ccacgatgtc acctcagccc cggacaacaa gccagccccg ggctccaccg ccccccagc 600
ccacggtgtc acctcggccc cggacaccag gccggccccg ggctccaccg ccccccagc 660
ccatggtgtc acctcggccc cggacaacag gcccgccttg ggctccaccg cccctccagt 720
ccacaatgtc acctcggcct caggctctgc atcaggctca gcttctactc tggtgcacaa 780
cggcacctct gccagggcta ccacaacccc agccagcaag agcactccat tctcaattcc 840
cagccaccac tctgatactc ctaccaccct tgccagccat agcaccaaga ctgatgccag 900
tagcactcac catagcacgg tacctcctct cacctcctcc aatcacagca cttctcccca 960
gttgtctact ggggtctctt tctttttcct gtcttttcac atttcaaacc tccagtttaa 1020
ttcctctctg gaagatccca gcaccgacta ctaccaagag ctgcagagag acatttctga 1080
aatgtttttg cagatttata aacaaggggg ttttctgggc ctctccaata ttaagttcag 1140
gccaggatct gtggtggtac aattgactct ggccttccga gaaggtacca tcaatgtcca 1200
cgacgtggag acacagttca atcagtataa aacggaagca gcctctcgat ataacctgac 1260
gatctcagac gtcagcgtga gtgatgtgcc atttcctttc tctgcccagt ctggggctgg 1320
ggtgccaggc tggggcatcg cgctgctggt gctggtctgt gttctggttg cgctggccat 1380
tgtctatctc attgccttgg ctgtctgtca gtgccgccga aagaactacg ggcagctgga 1440
catctttcca gcccgggata cctaccatcc tatgagcgag taccccacct accacaccca 1500
tgggcgctat gtgcccccta gcagtaccga tcgtagcccc tatgagaagg tttctgcagg 1560
taatggtggc agcagcctct cttacacaaa cccagcagtg gcagccactt ctgccaactt 1620
gtaggggcac gtcgcccgct gagctgagtg gccagccagt gccattccac tccactcagg 1680
ttcttcaggg ccagagcccc tgcaccctgt ttgggctggt gagctgggag ttcaggtggg 1740
ctgctcacag cctccttcag aggccccacc aatttctcgg acacttctca gtgtgtggaa 1800

gctcatgtgg gcccctgagg gctcatgcct gggaagtgtt gtggtggggg ctcccaggag 1860

gactggccca gagagccctg agatagcggg gatcctgaac tggactgaat aaaacgtggt 1920

ctcccactgc 1930

<210> 50
<211> 1798
<212> DNA
<213> Homo sapien

<400> 50
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctgtgagta tgaccagcag 360

cgtactctcc agccacagcc ccggttcagg ctcctccacc actcagggac aggatgtcac 420

tctggccccg gccacggaac cagcttcagg ttcagctgcc acctggggac aggatgtcac 480

ctcggtccca gtcaccaggc cagccctggg ctccaccacc ccgccagccc acgatgtcac 540

ctcggccccg gacaacaggc ccgccttggg ctccaccgcc cctccagtcc acaatgtcac 600

ctcggcctca ggctctgcat caggctcagc ttctactctg gtgcacaacg gcacctctgc 660

cagggctacc acaaccccag ccagcaagag cactccattc tcaattccca gccaccactc 720

tgatactcct accacccttg ccagccatag caccaagact gatgccagta gcactcacca 780

tagcacggta cctcctctca cctcctccaa tcacagcact tctccccagt tgtctactgg 840

ggtctctttc tttttcctgt cttttcacat ttcaaacctc cagtttaatt cctctctgga 900

agatcccagc accgactact accaagagct gcagagagac atttctgaaa tgtttttgca 960

gatttataaa caagggggtt ttctgggcct ctccaatatt aagttcaggc caggatctgt 1020

ggtggtacaa ttgactctgg ccttccgaga aggtaccatc aatgtccacg acgtggagac 1080

acagttcaat cagtataaaa cggaagcagc ctctcgatat aacctgacga tctcagacgt 1140

cagcgtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg tgccaggctg 1200

gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg tctatctcat 1260

tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca tctttccagc 1320

ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg ggcgctatgt 1380

gccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta atggtggcag 1440

cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt aggggcacgt 1500

cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt cttcagggcc 1560

agagcccctg caccctgttt gggctggtga gctgggagtt caggtgggct gctcacagcc 1620

tccttcagag gccccaccaa tttctcggac acttctcagt gtgtggaagc tcatgtgggc 1680

ccctgagggc tcatgcctgg gaagtgttgt ggtgggggct cccaggagga ctggcccaga 1740

gagccctgag atagcgggga tcctgaactg gactgaataa aacgtggtct cccactgc 1798

<210> 51
<211> 1312
<212> DNA
<213> Homo sapien

<400> 51
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctttgtcta ctggggtctc 360

tttctttttc ctgtcttttc acatttcaaa cctccagttt aattcctctc tggaagatcc 420

cagcaccgac tactaccaag agctgcagag agacatttct gaaatgtttt tgcagattta 480

taaacaaggg ggttttctgg gcctctccaa tattaagttc aggccaggat ctgtggtggt 540

acaattgact ctggccttcc gagaaggtac catcaatgtc cacgacgtgg agacacagtt 600

caatcagtat aaaacggaag cagcctctcg atataacctg acgatctcag acgtcagcgt 660

gagtgatgtg ccatttcctt tctctgccca gtctggggct ggggtgccag gctggggcat 720

cgcgctgctg gtgctggtct gtgttctggt tgcgctggcc attgtctatc tcattgcctt 780

ggctgtctgt cagtgccgcc gaaagaacta cgggcagctg gacatctttc cagcccggga 840

tacctaccat cctatgagcg agtaccccac ctaccacacc catgggcgct atgtgccccc 900

tagcagtacc gatcgtagcc cctatgagaa ggtttctgca ggtaatggtg gcagcagcct 960

ctcttacaca aacccagcag tggcagccac ttctgccaac ttgtaggggc acgtcgcccg 1020

ctgagctgag tggccagcca gtgccattcc actccactca ggttcttcag ggccagagcc 1080

cctgcaccct gtttgggctg gtgagctggg agttcaggtg ggctgctcac agcctccttc 1140

agaggcccca ccaatttctc ggacacttct cagtgtgtgg aagctcatgt gggcccctga 1200

gggctcatgc ctgggaagtg ttgtggtggg ggctcccagg aggactggcc cagagagccc 1260

tgagatagcg gggatcctga actggactga ataaaacgtg gtctcccact gc 1312

<210> 52
<211> 2094
<212> DNA
<213> Homo sapien

<400> 52
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggtaca gttctgggtg tggacccagt gtggtggttg gagggtgggt 1200

ggtggtcatg accgtaggga gggactggtg cacttaaggt tgggggaaga gtgctgagcc 1260

agagctggga cccgtggctg aagtgcccat ttccctgtga ccaggccagg atctgtggtg 1320

gtacaattga ctctggcctt ccgagaaggt accatcaatg tccacgacgt ggagacacag 1380

ttcaatcagt ataaaacgga agcagcctct cgatataacc tgacgatctc agacgtcagc 1440

gtgagtgatg tgccatttcc tttctctgcc cagtctgggg ctggggtgcc aggctggggc 1500

atcgcgctgc tggtgctggt ctgtgttctg gttgcgctgg ccattgtcta tctcattgcc 1560

ttggctgtct gtcagtgccg ccgaaagaac tacgggcagc tggacatctt tccagcccgg 1620

gatacctacc atcctatgag cgagtacccc acctaccaca cccatgggcg ctatgtgccc 1680

cctagcagta ccgatcgtag cccctatgag aaggtttctg caggtaatgg tggcagcagc 1740

ctctcttaca caaacccagc agtggcagcc acttctgcca acttgtaggg gcacgtcgcc 1800

cgctgagctg agtggccagc cagtgccatt ccactccact caggttcttc agggccagag 1860

cccctgcacc ctgtttgggc tggtgagctg ggagttcagg tgggctgctc acagcctcct 1920

tcagaggccc caccaatttc tcggacactt ctcagtgtgt ggaagctcat gtgggcccct 1980

gagggctcat gcctgggaag tgttgtggtg gggctccca ggaggactgg cccagagagc 2040

cctgagatag cggggatcct gaactggact gaataaaacg tggtctccca ctgc 2094

<210> 53
<211> 2049
<212> DNA
<213> Homo sapien

<400> 53
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cgtgagtgat gtgccatttc ctttctctgc 1320

ccagtctggg gctggggtgc caggctgggg catcgcgctg ctggtgctgg tctgtgttct 1380

ggttgcgctg gccattgtct atctcattgc cttggctgtc tgtcagtgcc gccgaaagaa 1440

ctacgggcag ctggacatct ttccagcccg ggatacctac catcctatga gcgagtaccc 1500

cacctaccac acccatgggc gctatgtgcc ccctagcagt accgatcgta gcccctatga 1560

gaaggtttct gcaggtaatg gtggcagcag cctctcttac acaaacccag cagtggcagc 1620

cacttctgcc aacttgtagg ggcacgtcgc ccgctgagct gagtggccag ccagtgccat 1680

tccactccac tcaggttctt cagggccaga gcccctgcac cctgtttggg ctggtgagct 1740

gggagttcag gtgggctgct cacagcctcc ttcagaggcc ccaccaattt ctcggacact 1800

tctcagtgtg tggaagctca tgtgggcccc tgagggctca tgcctgggaa gtgttgtggt 1860

gggggctccc aggaggactg gcccagagag ccctgagata gcggggatcc tgaactggac 1920

tgaataaaac gtggtctccc actgcaaaag acataaaaaa agaaaaagac aaagacgagc 1980

aaaaagacaa aaagaggcaa aaacaacaaa acacaacaaa caaaaaaaag cacacacaaa 2040

aaaaagaag 2049

<210> 54
<211> 2194
<212> DNA
<213> Homo sapien

<400> 54
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660
caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720
caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780
tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840
tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900
caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960
cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020
aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080
gagagacatt tctgaaatgg tgagtatcgg cctttccttc cccatgctcc cctgaagcag 1140
ccatcagaac tgtccacacc ctttgcatca agcctgagtc ctttccctct caccccagtt 1200
ttttgcagat ttataaacaa gggggttttc tgggcctctc caatattaag ttcaggtaca 1260
gttctgggtg tggacccagt gtggtggttg gagggtgggt ggtggtcatg accgtaggga 1320
gggactggtg cacttaaggt tgggggaaga gtgctgagcc agagctggga cccgtggctg 1380
aagtgcccat ttccctgtga ccaggccagg atctgtggtg gtacaattga ctctggcctt 1440
ccgagaaggt accatcaatg tccacgacgt ggagacacag ttcaatcagt ataaaacgga 1500
agcagcctct cgatataacc tgacgatctc agacgtcagc gtgagtgatg tgccatttcc 1560
tttctctgcc cagtctgggg ctggggtgcc aggctggggc atcgcgctgc tggtgctggt 1620
ctgtgttctg gttgcgctgg ccattgtcta tctcattgcc ttggctgtct gtcagtgccg 1680
ccgaaagaac tacgggcagc tggacatctt tccagcccgg gatacctacc atcctatgag 1740
cgagtacccc acctaccaca cccatgggcg ctatgtgccc cctagcagta ccgatcgtag 1800
cccctatgag aaggtttctg caggtaatgg tggcagcagc ctctcttaca caaacccagc 1860
agtggcagcc acttctgcca acttgtaggg gcacgtcgcc cgctgagctg agtggccagc 1920
cagtgccatt ccactccact caggttcttc agggccagag cccctgcacc ctgtttgggc 1980
tggtgagctg ggagttcagg tgggctgctc acagcctcct tcagaggccc caccaatttc 2040
tcggacactt ctcagtgtgt ggaagctcat gtgggcccct gagggctcat gcctgggaag 2100
tgttgtggtg gggctcccca ggaggactgg cccagagagc cctgagatag cggggatcct 2160
gaactggact gaataaaacg tggtctccca ctgc 2194

<210> 55
<211> 1183
<212> DNA
<213> Homo sapien

<400> 55
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctatcccag caccgactac 360

taccaagagc tgcagagaga catttctgaa atggccagga tctgtggtgg tacaattgac 420

tctggccttc cgagaaggta ccatcaatgt ccacgacgtg gagacacagt tcaatcagta 480

taaaacggaa gcagcctctc gatataacct gacgatctca gacgtcagcg tgagtgatgt 540

gccatttcct ttctctgccc agtctggggc tggggtgcca ggctggggca tcgcgctgct 600

ggtgctggtc tgtgttctgg ttgcgctggc cattgtctat ctcattgcct tggctgtctg 660

tcagtgccgc cgaaagaact acgggcagct ggacatcttt ccagcccggg atacctacca 720

tcctatgagc gagtacccca cctaccacac ccatgggcgc tatgtgcccc ctagcagtac 780

cgatcgtagc ccctatgaga aggtttctgc aggtaatggt ggcagcagcc tctcttacac 840

aaacccagca gtggcagcca cttctgccaa cttgtagggg cacgtcgccc gctgagctga 900

gtggccagcc agtgccattc cactccactc aggttcttca gggccagagc ccctgcaccc 960

tgtttgggct ggtgagctgg gagttcaggt gggctgctca cagcctcctt cagaggcccc 1020

accaatttct cggacacttc tcagtgtgtg gaagctcatg tgggcccctg agggctcatg 1080

cctgggaagt gttgtggtgg gggctcccag gaggactggc ccagagagcc ctgagatagc 1140

ggggatcctg aactggactg aataaaacgt ggtctcccac tgc 1183

<210> 56
<211> 2333
<212> DNA
<213> Homo sapien

<400> 56
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cgtgctgtga ttggaggagg tgagaggagg 1320

taccgtgcta tggtgagtgc tactggcatc agtcttggtg ctatggctgg caagggtggt 1380

aggagtatca gagtggtggc tgggaattga gaatggagtg ctcttgctgg ctggggttgt 1440

ggtagccctg gcagaggtgc cgttgtgcac cagagtagaa gctgagcctg atgccagtag 1500

cactcaccat agcacggtac ctcctctcac ctcctccaat cacagcactt ctccccagtt 1560

gtctactggg gtctctttct ttttcctgtc ttttcaattt caaacctcca gtttaattcc 1620

tctctggaag atcccagcac cgactactac caagagctgc agagagacat ttctgaaatg 1680

tgagtgatgt gccatttcct ttctctgccc agtctggggc tggggtgcca ggctggggca 1740

tcgcgctgct ggtgctggtc tgtgttctgg ttgcgctggc cattgtctat ctcattgcct 1800

tggctgtctg tcagtgccgc cgaaagaact acgggcagct ggacatcttt ccagcccggg 1860

atacctacca tcctatgagc gagtacccca cctaccacac ccatgggcgc tatgtgcccc 1920

ctagcagtac cgatcgtagc ccctatgaga aggtttctgc aggtaatggt ggcagcagcc 1980

tctcttacac aaacccagca gtggcagcca cttctgccaa cttgtagggg cacgtcgccc 2040

gctgagctga gtggccagcc agtgccattc cactccactc aggttcttca gggccagagc 2100

ccctgcaccc tgtttgggct ggtgagctgg gagttcaggt gggctgctca cagcctcctt 2160

cagaggcccc accaatttct cggacacttc tcagtgtgtg gaagctcatg tgggcccctg 2220

agggctcatg cctgggaagt gttgtggtgg gggctcccag gaggactggc ccagagagcc 2280

ctgagatagc ggggatcctg aactggactg aataaaacgt ggtctcccac tgc 2333

<210> 57
<211> 1712
<212> DNA
<213> Homo sapien

<400> 57
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ggggtgccag gctggggcat cgcgctgctg gtgctggtct 1140

gtgttctggt tgcgctggcc attgtctatc tcattgcctt ggctgtctgt cagtgccgcc 1200

gaaagaacta cgggcagctg gacatctttc cagcccggga tacctaccat cctatgagcg 1260

agtaccccac ctaccacacc catgggcgct atgtgccccc tagcagtacc gatcgtagcc 1320

cctatgagaa ggtttctgca ggtaatggtg gcagcagcct ctcttacaca aacccagcag 1380

tggcagccac ttctgccaac ttgtaggggc acgtcgcccg ctgagctgag tggccagcca 1440

gtgccattcc actccactca ggttcttcag ggccagagcc cctgcaccct gtttgggctg 1500

gtgagctggg agttcaggtg ggctgctcac agcctccttc agaggcccca ccaatttctc 1560

ggacacttct cagtgtgtgg aagctcatgt gggcccctga gggctcatgc ctgggaagtg 1620

ttgtggtggg ggctcccagg aggactggcc cagagagccc tgagatagcg gggatcctga 1680

actggactga ataaaacgtg gtctcccact gc 1712

<210> 58
<211> 1605
<212> DNA
<213> Homo sapien

<400> 58
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctgtgagta tgaccagcag 360

cgtactctcc agccacagcc ccggttcagg ctcctccacc actcagggac aggatgtcac 420

tctggccccg gccacggaac cagcttcagg ttcagctgcc acctggggac aggatgtcac 480

ctcggtccca gtcaccaggc cagccctggg ctccaccacc ccgccagccc acgatgtcac 540

ctcggccccg gacaacaggc ccgccttggg ctccaccgcc cctccagtcc acaatgtcac 600

ctcggcctca ggctctgcat caggctcagc ttctactctg gtgcacaacg gcacctctgc 660

cagggctacc acaaccccag ccagcaagag cactccattc tcaattccca gccaccactc 720

tgatactcct accacccttg ccagccatag caccaagact gatgccagta gcactcacca 780

tagcacggta cctcctctca cctcctccaa tcacagcact tctccccagt tgtctactgg 840

ggtctctttc tttttcctgt cttttcacat ttcaaacctc cagtttaatt cctctctgga 900

agatcccagc accgactact accaagagct gcagagagac atttctgaaa tgtgagtgat 960

gtgccatttc ctttctctgc ccagtctggg gctggggtgc caggctgggg catcgcgctg 1020

ctggtgctgg tctgtgttct ggttgcgctg gccattgtct atctcattgc cttggctgtc 1080

tgtcagtgcc gccgaaagaa ctacgggcag ctggacatct ttccagcccg ggatacctac 1140

catcctatga gcgagtaccc cacctaccac acccatgggc gctatgtgcc ccctagcagt 1200

accgatcgta gcccctatga gaaggtttct gcaggtaatg gtggcagcag cctctcttac 1260

acaaacccag cagtggcagc cacttctgcc aacttgtagg ggcacgtcgc ccgctgagct 1320

gagtggccag ccagtgccat tccactccac tcaggttctt cagggccaga gcccctgcac 1380

cctgtttggg ctggtgagct gggagttcag gtgggctgct cacagcctcc ttcagaggcc 1440

ccaccaattt ctcggacact tctcagtgtg tggaagctca tgtgggcccc tgagggctca 1500

tgcctgggaa gtgttgtggt gggggctccc aggaggactg gcccagagag ccctgagata 1560

gcggggatcc tgaactggac tgaataaaac gtggtctccc actgc 1605

<210> 59
<211> 1874
<212> DNA
<213> Homo sapien

<400> 59
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cgtgagtgat gtgccatttc ctttctctgc 1320

ccagtctggg gctggggtgc caggctgggg catcgcgctg ctggtgctgg tctgtgttct 1380

ggttgcgctg gccattgtct atctcattgc cttggctgtc tgtcagtgcc gccgaaagaa 1440

ctacgggcag ctggacatct ttccagcccg ggatacctac catcctatga gcgagtaccc 1500

cacctaccac acccatgggc gctatgtgcc ccctagcagt accgatcgta gcccctatga 1560

gaaggtgaga ttgggcccca caggccaggg gaagcagagg gtttggctgg gcaaggattc 1620

tgaagggggt acttggaaaa cccaaagagc ttggaagagg tgagaagtgg cgtgaagtga 1680

gcaggggagg gcctggcaag gatgaggggc agaggtcaga ggagttttgg gggacaggcc 1740

tgggaggaga ctatggaaga aaggggccct caagagggag tggccccact gccagaattc 1800

ctaaaagatc attggccgtc cacattcatg ctggctggcg ctggctgaac tggtgccacc 1860

gtggcagttt tgtt 1874

<210> 60
<211> 1634
<212> DNA
<213> Homo sapien

<400> 60
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cgtgagtgat gtgccatttc ctttctctgc 1320

ccagtctggg gctggggtgc caggctgggg catcgcgctg ctggtgctgg tctgtgttct 1380

ggttgcgctg gccattgtct atctcattgc cttggctgtc tgtcagtgcc gccgaaagaa 1440

ctacgggcag ctggacatct ttccagcccg ggatacctac catcctatga gcgagtggag 1500

ggtgtagaag agaagaagaa ggaggttcct gctgtgccag aaacccttaa gaaaaagcga 1560

aggaatttcg cagagctgaa gatcaagcgc ctgagaaaga agttksccaa aagatgcttc 1620

gaaaggcaag gagg 1634

<210> 61
<211> 943
<212> DNA
<213> Homo sapien

<400> 61
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gcttttaatt cctctctgga 360

agatcccagc accgactact accaagagct gcagagagac atttctgaaa tggctgtctg 420

tcagtgccgc cgaaagaact acgggcagct ggacatcttt ccagcccggg atacctacca 480

tcctatgagc gagtacccca cctaccacac ccatgggcgc tatgtgcccc ctagcagtac 540

cgatcgtagc ccctatgaga aggtttctgc aggtaatggt ggcagcagcc tctcttacac 600

aaacccagca gtggcagcca cttctgccaa cttgtagggg cacgtcgccc gctgagctga 660

gtggccagcc agtgccattc cactccactc aggttcttca gggccagagc ccctgcaccc 720

tgtttgggct ggtgagctgg gagttcaggt gggctgctca cagcctcctt cagaggcccc 780

accaatttct cggacacttc tcagtgtgtg gaagctcatg tgggcccctg agggctcatg 840

cctgggaagt gttgtggtgg gggctcccag gaggactggc ccagagagcc ctgagatagc 900

ggggatcctg aactggactg aataaaacgt ggtctcccac tgc 943

<210> 62

<211> 997
<212> DNA
<213> Homo sapien

<400> 62
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctttgtcta ctggggtctc 360

tttctttttc ctgtcttttc acatttcaaa cctccagttt aattcctctc tggaagatcc 420

cagcaccgac tactaccaag agctgcagag agacatttct gaaatggctg tctgtcagtg 480

ccgccgaaag aactacgggc agctggacat ctttccagcc cgggatacct accatcctat 540

gagcgagtac cccacctacc acacccatgg gcgctatgtg ccccctagca gtaccgatcg 600

tagcccctat gagaaggttt ctgcaggtaa tggtggcagc agcctctctt acacaaaccc 660

agcagtggca gccacttctg ccaacttgta ggggcacgtc gcccgctgag ctgagtggcc 720

agccagtgcc attccactcc actcaggttc ttcagggcca gagcccctgc accctgtttg 780

ggctggtgag ctgggagttc aggtgggctg ctcacagcct ccttcagagg ccccaccaat 840

ttctcggaca cttctcagtg tgtggaagct catgtgggcc cctgagggct catgcctggg 900

aagtgttgtg gtgggggctc ccaggaggac tggcccagag agccctgaga tagcggggat 960

cctgaactgg actgaataaa acgtggtctc ccactgc 997

<210> 63
<211> 548
<212> DNA
<213> Homo sapien

<400> 63
gaagcccgga atggcttacc ttgatcagca gccctggtaa gaactacggg cagctggaca 60

tctttccagc ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg 120

ggcgctatgt gccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta 180

atggtggcag cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt 240

aggggcacgt cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt 300

cttcagggcc agagcccctg caccctgttt gggctggtga gctgggagtt caggtgggct 360

gctcacagcc tccttcagag gccccaccaa tttctcggac acttctcagt gtgtggaagc 420

tcatgtgggc ccctgagggc tcatgcctgg gaagtgttgt ggtgggggct cccaggagga 480

ctggcccaga gagccctgag atagcgggga tcctgaactg gactgaataa aacgtggtct 540

cccactgc 548

<210> 64
<211> 1378
<212> DNA
<213> Homo sapien

<400> 64
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc accettgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cgctgaagta ccatttcaca tcatgctgac 1320

caatatgggc ccatggagta ccacaacgtc ggggcaatcc gatttcggca caactact 1378

<210> 65
<211> 162

<212> DNA
<213> Homo sapien

<400> 65
gcggccgcct actactacta ctgctcgaat tcaagcttct aacgatgtac gggctcatgc 60

ctgggaagtg ttgtggtggg ggctcccagg aggactggcc cagagagccc tgagatagcg 120

gggatcctga actggactga ataaaacgtg gtctcccact gc 162

<210> 66
<211> 1285
<212> DNA
<213> Homo sapien

<400> 66
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 420

gagagacatt tctgaaatgt tttgcagat ttataaacaa gggggttttc tgggcctctc 480

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 540

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 600

tcgatataac ctgacgatct cagacgtcag cgtgagtgat gtgccatttc ctttctctgc 660

ccagtctggg gctggggtgc caggctgggg catcgcgctg ctggtgctgg tctgtgttct 720

ggttgcgctg gccattgtct atctcattgc cttggctgtc tgtcagtgcc gccgaaagaa 780

ctacgggcag ctggacatct ttccagcccg ggatacctac catcctatga gcgagtaccc 840

cacctaccac acccatgggc gctatgtgcc ccctagcagt accgatcgta gcccctatga 900

gaaggtttct gcaggtaatg gtggcagcag cctctcttac acaaacccag cagtggcagc 960

cacttctgcc aacttgtagg ggcacgtcgc ccgctgagct gagtggccag ccagtgccat 1020

tccactccac tcaggttctt cagggccaga gcccctgcac cctgtttggg ctggtgagct 1080

gggagttcag gtgggctgct cacagcctcc ttcagaggcc ccaccaattt ctcggacact 1140

tctcagtgtg tggaagctca tgtgggcccc tgagggctca tgcctgggaa gtgttgtggt 1200

gggggctccc aggaggactg gcccagagag ccctgagata gcggggatcc tgaactggac 1260

tgaataaaac gtggtctccc actgc 1285

<210> 67
<211> 1517
<212> DNA
<213> Homo sapien

<400> 67
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct tttttgcaga tttataaaca agggggtttt ctgggcctct ccaatattaa 420

gttcaggcca ggatctgtgg tggtacaatt gactctggcc ttccgagaag gtaccatcaa 480

tgtccacgac gtggagacac agttcaatca gtataaaacg gaagcagcct ctcgatataa 540

cctgacgatc tcagacgtca gcgtgagtga tgtgccattt cctttctctg cccagtctgg 600

ggctggggtg ccaggctggg gcatcgcgct gctggtgctg gtctgtgttc tggttgcgct 660

ggccattgtc tatctcattg ccttggctgt ctgtcagtgc cgccgaaaga actacgggca 720

gctggacatc tttccagccc gggataccta ccatcctatg agcgagtacc ccacctacca 780

cacccatggg cgctatgtgc cccctagcag taccgatcgt agcccctatg agaaggtttc 840

tgcaggtaat ggtggcagca gcctctctta cacaaaccca gcagtggcag ccacttctgc 900

caacttgtag gggcacgtcg cccgctgagc tgagtggcca gccagtgcca ttccactcca 960

ctcaggttct tcagggccag agcccctgca ccctgtttgg gctggtgagc tgggagttca 1020

ggtgggctgc tcacagcctc cttcagaggc cccaccaatt tctcggacac ttctcagtgt 1080

gtggaagctc atgtgggccc ctgagggctc atgcctggga agtgttgtgg tgggggctcc 1140

caggaggact ggcccagaga gccctgagat agcggggatc ctgaactgga ctgaataaaa 1200

cgtggtctcc cactgcaaaa aaaaaagaag actgagaagc ggtcgtaaaa ggagcgcacg 1260

cagaggcggc tggagggcga tgacactagt gcgaactaga gacgggagag agagtgggca 1320

cgagccgata gataggtgtg gtggtgcgga gtcgctgtgc gggcgatggg cgggcacggg 1380

ggatgtgtcc tacgaccgga gcggtcggta gccgccatgg cagtgtggag tcgcggagta 1440

cagtcgactg gggcgactca cacgaacgta catgtacacg tgtacacgca agctacgtgt 1500

gtgagcggca gagattg 1517

<210> 68

<211> 524
<212> DNA
<213> Homo sapien

<400> 68
gccctgatca gagcccccg gtagaaggca ctccatggcc tgccataacc tcctatctcc 60

ccaggctgtc tgtcagtgcc gccgaaagaa ctacgggcag ctggacatct ttccagcccg 120

ggatacctac catcctatga gcgagtaccc cacctaccac acccatgggc gctatgtgcc 180

ccctagcagt accgatcgta gcccctatga gaaggtgaga ttgggcccca caggccaggg 240

gaagcagagg gtttggctgg gcaaggattc tgaagggggt acttggaaaa cccaaagagc 300

ttggaagagg tgagaagtgg cgtgaagtga gcaggggagg gcctggcaag gatgaggggc 360

agaggtcaga ggagttttgg gggacaggcc tgggaggaga ctatggaaga aaggggccct 420

caagagggag tggccccact gccagaattc ctaaaagatc attggccgtc cacattcatg 480

ctggctggcg ctggctgaac tggtgccacc gtggcagttt tgtt 524

<210> 69
<211> 1949
<212> DNA
<213> Homo sapien

<400> 69
agggggaaga gagtagggag agggaaggct taagagggga agaaatgcag gggccatgag 60

ccaaggccta tgggcagaga gaaggaggct gctgcaggga aggaggcggc caacccaggg 120

gttactgagg ctgcccactc cccagtcctc ctggtattat ttctctggtg gccagagctt 180

atattttctt cttgctctta tttttccttc ataaagaccc aaccctatga ctttaacttc 240

ttacagctac cacagcccct aaacccgcaa cagttgttac aggttctggt catgcaagct 300

ctaccccagg tggagaaaag gagacttcgg ctacccagag aagttcagtg cccagctcta 360

ctgagaagaa tgctgtgagt atgaccagca gcgtactctc cagccacagc cccggttcag 420

gctcctccac cactcaggga caggatgtca ctctggcccc ggccacggaa ccagcttcag 480

gttcagctgc cacctgggga caggatgtca cctcggtccc agtcaccagg ccagccctgg 540

gctccaccac cccgccagcc cacgatgtca cctcagcccc ggacaacaag ccagccccgg 600

gctccaccgc ccccccagcc cacggtgtca cctcggcccc ggacaccagg ccggccccgg 660

gctccaccgc ccccccagcc catggtgtca cctcggcccc ggacaacagg cccgccttgg 720

gctccaccgc ccctccagtc cacaatgtca cctcggcctc aggctctgca tcaggctcag 780

cttctactct ggtgcacaac ggcacctctg ccagggctac cacaacccca gccagcaaga 840

gcactccatt ctcaattccc agccaccact ctgatactcc taccaccctt gccagccata 900

gcaccaagac tgatgccagt agcactcacc atagcacggt acctcctctc acctcctcca 960

atcacagcac ttctccccag ttgtctactg gggtctcttt ctttttcctg tctttttcaca 1020

tttcaaacct ccagtttaat tcctctctgg aagatcccag caccgactac taccaagagc 1080

tgcagagaga catttctgaa atgtttttgc agatttataa acaagggggt tttctgggcc 1140

tctccaatat taagttcagg ccaggatctg tggtggtaca attgactctg gccttccgag 1200

aaggtaccat caatgtccac gacgtggaga cacagttcaa tcagtataaa acggaagcag 1260

cctctcgata taacctgacg atctcagacg tcagcgtgag tgatgtgcca tttcctttct 1320

ctgcccagtc tggggctggg gtgccaggct ggggcatcgc gctgctggtg ctggtctgtg 1380

ttctggttgc gctggccatt gtctatctca ttgccttggc tgtctgtcag tgccgccgaa 1440

agaactacgg gcagctggac atctttccag cccgggatac ctaccatcct atgagcgagt 1500

accccaccta ccacacccat gggcgctatg tgccccctag cagtaccgat cgtagcccct 1560

atgagaaggt ttctgcaggt aatggtggca gcagcctctc ttacacaaac ccagcagtgg 1620

cagccacttc tgccaacttg taggggcacg tcgcccgctg agctgagtgg ccagccagtg 1680

ccattccact ccactcaggt tcttcagggc cagagcccct gcaccctgtt tgggctggtg 1740

agctgggagt tcaggtgggc tgctcacagc ctccttcaga ggccccacca atttctcgga 1800

cacttctcag tgtgtggaag ctcatgtggg cccctgaggg ctcatgcctg ggaagtgttg 1860

tggtgggggc tcccaggagg actggcccag agagccctga gatagcgggg atcctgaact 1920

ggactgaata aaacgtggtc tcccactgc 1949

<210> 70
<211> 1803
<212> DNA
<213> Homo sapien

<400> 70
ggtagcgcaa gcagaacaca gaccagcacc agcagcgcga tgccccagcc gggcacccag 60

tctcctttct tcctgctgct gctcctcaca gtgcttacag ctaccacagc ccctaaaccc 120

gcaacagttg ttacaggttc tggtcatgca agctctaccc caggtggaga aaaggagact 180

tcggctaccc agagaagttc agtgcccagc tctactgaga agaatgctgt gagtatgacc 240

agcagcgtac tctccagcca cagccccggt tcaggctcct ccaccactca gggacaggat 300

gtcactctgg ccccggccac ggaaccagct tcaggttcag ctgccacctg gggacaggat 360

gtcacctcgg tcccagtcac caggccagcc ctgggctcca ccaccccgcc agcccacgat 420

gtcacctcag ccccggacaa caagccagcc ccgggctcca ccgccccccc agcccacggt 480

gtcacctcgg ccccggacac caggccggcc ccgggctcca ccgccccccc agcccatggt 540

gtcacctcgg ccccggacaa caggcccgcc ttgggctcca ccgcccctcc agtccacaat 600

```
gtcacctcgg cctcaggctc tgcatcaggc tcagcttcta ctctggtgca caacggcacc    660

tctgccaggg ctaccacaac cccagccagc aagagcactc cattctcaat tcccagccac    720

cactctgata ctcctaccac ccttgccagc catagcacca agactgatgc cagtagcact    780

caccatagca cggtacctcc tctcacctcc tccaatcaca gcacttctcc ccagttgtct    840

actggggtct ctttcttttt cctgtctttt cacatttcaa acctccagtt taattcctct    900

ctggaagatc ccagcaccga ctactaccaa gagctgcaga gagacatttc tgaaatgttt    960

ttgcagattt ataaacaagg gggttttctg ggcctctcca atattaagtt caggccagga   1020

tctgtggtgg tacaattgac tctggccttc cgagaaggta ccatcaatgt ccacgacgtg   1080

gagacacagt tcaatcagta taaaacggaa gcagcctctc gatataacct gacgatctca   1140

gacgtcagcg tgagtgatgt gccatttcct ttctctgccc agtctggggc tggggtgcca   1200

ggctggggca tcgcgctgct ggtgctggtc tgtgttctgg ttgcgctggc cattgtctat   1260

ctcattgcct tggctgtctg tcagtgccgc cgaaagaact acgggcagct ggacatcttt   1320

ccagcccggg atacctacca tcctatgagc gagtacccca cctaccacac ccatgggcgc   1380

tatgtgcccc ctagcagtac cgatcgtagc ccctatgaga aggtttctgc aggtaatggt   1440

ggcagcagcc tctcttacac aaacccagca gtggcagcca cttctgccaa cttgtagggg   1500

cacgtcgccc gctgagctga gtggccagcc agtgccattc cactccactc aggttcttca   1560

gggccagagc ccctgcaccc tgtttgggct ggtgagctgg gagttcaggt gggctgctca   1620

cagcctcctt cagaggcccc accaatttct cggacacttc tcagtgtgtg gaagctcatg   1680

tgggcccctg agggctcatg cctgggaagt gttgtggtgg gggctcccag gaggactggc   1740

ccagagagcc ctgagatagc ggggatcctg aactggactg aataaaacgt ggtctcccac   1800

tgc                                                                 1803
```

<210> 71
<211> 1258
<212> DNA
<213> Homo sapien

<400> 71

```
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg     60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac    120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac    180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac    240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc    300

tacccagaga agttcagtgc ccagctctac tgagaagaat gcttttaatt cctctctgga    360
```

agatcccagc accgactact accaagagct gcagagagac atttctgaaa tgtttttgca 420

gatttataaa caaggggtt ttctgggcct ctccaatatt aagttcaggc caggatctgt 480

ggtggtacaa ttgactctgg ccttccgaga aggtaccatc aatgtccacg acgtggagac 540

acagttcaat cagtataaaa cggaagcagc ctctcgatat aacctgacga tctcagacgt 600

cagcgtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg tgccaggctg 660

gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg tctatctcat 720

tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca tctttccagc 780

ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg ggcgctatgt 840

gcccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta atggtggcag 900

cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt aggggcacgt 960

cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt cttcagggcc 1020

agagcccctg caccctgttt gggctggtga gctgggagtt caggtgggct gctcacagcc 1080

tccttcagag gccccaccaa tttctcggac acttctcagt gtgtggaagc tcatgtgggc 1140

ccctgagggc tcatgcctgg gaagtgttgt ggtgggggct cccaggagga ctggcccaga 1200

gagccctgag atagcgggga tcctgaactg gactgaataa aacgtggtct cccactgc 1258

<210> 72
<211> 2045
<212> DNA
<213> Homo sapien

<400> 72
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc accettgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgg tgagtatcgg cctttccttc cccatgctcc cctgaagcag 1140

ccatcagaac tgtccacacc ctttgcatca agcctgagtc ctttccctct caccccagtt 1200

ttttgcagat ttataaacaa gggggttttc tgggcctctc caatattaag ttcaggccag 1260

gatctgtggt ggtacaattg actctggcct tccgagaagg taccatcaat gtccacgacg 1320

tggagacaca gttcaatcag tataaaacgg aagcagcctc tcgatataac ctgacgatct 1380

cagacgtcag cgtgagtgat gtgccatttc ctttctctgc ccagtctggg gctggggtgc 1440

caggctgggg catcgcgctg ctggtgctgg tctgtgttct ggttgcgctg gccattgtct 1500

atctcattgc cttggctgtc tgtcagtgcc gccgaaagaa ctacgggcag ctggacatct 1560

ttccagcccg ggatacctac catcctatga gcgagtaccc cacctaccac acccatgggc 1620

gctatgtgcc ccctagcagt accgatcgta gcccctatga gaaggtttct gcaggtaatg 1680

gtggcagcag cctctcttac acaaacccag cagtggcagc cacttctgcc aacttgtagg 1740

ggcacgtcgc ccgctgagct gagtggccag ccagtgccat tccactccac tcaggttctt 1800

cagggccaga gcccctgcac cctgtttggg ctggtgagct gggagttcag gtgggctgct 1860

cacagcctcc ttcagaggcc ccaccaattt ctcggacact tctcagtgtg tggaagctca 1920

tgtgggcccc tgagggctca tgcctgggaa gtgttgtggt gggggctccc aggaggactg 1980

gcccagagag ccctgagata gcggggatcc tgaactggac tgaataaaac gtggtctccc 2040

actgc 2045

<210> 73
<211> 1266
<212> DNA
<213> Homo sapien

<400> 73
taggaggtag ggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct atcccagcac cgactactac caagagctgc agagagacat ttctgaaatg 420

tttttgcaga tttataaaca agggggtttt ctgggcctct ccaatattaa gttcaggcca 480

ggatctgtgg tggtacaatt gactctggcc ttccgagaag gtaccatcaa tgtccacgac 540

gtggagacac agttcaatca gtataaaacg gaagcagcct ctcgatataa cctgacgatc 600

tcagacgtca gcgtgagtga tgtgccattt cctttctctg cccagtctgg ggctggggtg 660

ccaggctggg gcatcgcgct gctggtgctg gtctgtgttc tggttgcgct ggccattgtc 720

tatctcattg ccttggctgt ctgtcagtgc cgccgaaaga actacgggca gctggacatc 780

tttccagccc gggataccta ccatcctatg agcgagtacc ccacctacca cacccatggg 840

cgctatgtgc cccctagcag taccgatcgt agcccctatg agaaggtttc tgcaggtaat 900

ggtggcagca gcctctctta cacaaaccca gcagtggcag ccacttctgc caacttgtag 960

gggcacgtcg cccgctgagc tgagtggcca gccagtgcca ttccactcca ctcaggttct 1020

tcagggccag agcccctgca ccctgtttgg gctggtgagc tgggagttca ggtgggctgc 1080

tcacagcctc cttcagaggc cccaccaatt tctcggacac ttctcagtgt gtggaagctc 1140

atgtgggccc ctgagggctc atgcctggga agtgttgtgg tgggggctcc caggaggact 1200

ggcccagaga gccctgagat agcggggatc ctgaactgga ctgaataaaa cgtggtctcc 1260

cactgc 1266

<210> 74
<211> 1189
<212> DNA
<213> Homo sapien

<400> 74
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctttttgc agatttataa 360

acaagggggt ttctgggcc tctccaatat taagttcagg ccaggatctg tggtggtaca 420

attgactctg gccttccgag aaggtaccat caatgtccac gacgtggaga cacagttcaa 480

tcagtataaa acggaagcag cctctcgata taacctgacg atctcagacg tcagcgtgag 540

tgatgtgcca tttcctttct ctgcccagtc tggggctggg gtgccaggct ggggcatcgc 600

gctgctggtg ctggtctgtg ttctggttgc gctggccatt gtctatctca ttgccttggc 660

tgtctgtcag tgccgccgaa agaactacgg gcagctggac atctttccag cccgggatac 720

ctaccatcct atgagcgagt accccaccta ccacacccat gggcgctatg tgccccctag 780

cagtaccgat cgtagcccct atgagaaggt ttctgcaggt aatggtggca gcagcctctc 840

ttacacaaac ccagcagtgg cagccacttc tgccaacttg taggggcacg tcgcccgctg 900

agctgagtgg ccagccagtg ccattccact ccactcaggt tcttcagggc cagagcccct 960

gcaccctgtt tgggctggtg agctgggagt tcaggtgggc tgctcacagc ctccttcaga 1020

ggccccacca atttctcgga cacttctcag tgtgtggaag ctcatgtggg cccctgaggg 1080

ctcatgcctg ggaagtgttg tggtgggggc tcccaggagg actggcccag agagccctga 1140

gatagcgggg atcctgaact ggactgaata aaacgtggtc tcccactgc 1189

<210> 75
<211> 1216
<212> DNA
<213> Homo sapien

<400> 75
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct ttttgcaga tttataaaca agggggtttt ctgggcctct ccaatattaa 420

gttcaggcca ggatctgtgg tggtacaatt gactctggcc ttccgagaag gtaccatcaa 480

tgtccacgac gtggagacac agttcaatca gtataaaacg gaagcagcct ctcgatataa 540

cctgacgatc tcagacgtca gcgtgagtga tgtgccattt cctttctctg cccagtctgg 600

ggctggggtg ccaggctggg gcatcgcgct gctggtgctg gtctgtgttc tggttgcgct 660

ggccattgtc tatctcattg ccttggctgt ctgtcagtgc cgccgaaaga actacgggca 720

gctggacatc tttccagccc gggataccta ccatcctatg agcgagtacc ccacctacca 780

cacccatggg cgctatgtgc cccctagcag taccgatcgt agcccctatg agaaggtttc 840

tgcaggtaat ggtggcagca gcctctctta cacaaaccca gcagtggcag ccacttctgc 900

caacttgtag ggcacgtcg cccgctgagc tgagtggcca gccagtgcca ttccactcca 960

ctcaggttct tcagggccag agcccctgca ccctgtttgg gctggtgagc tgggagttca 1020

ggtgggctgc tcacagcctc cttcagaggc cccaccaatt tctcggacac ttctcagtgt 1080

gtggaagctc atgtgggccc ctgagggctc atgcctggga agtgttgtgg tgggggctcc 1140

caggaggact ggcccagaga gccctgagat agcggggatc ctgaactgga ctgaataaaa 1200

cgtggtctcc cactgc 1216

<210> 76
<211> 2090
<212> DNA
<213> Homo sapien

<400> 76
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cggtgaggct acttccctgg ctgcagccca 1320

gcaccatgcc ggggcccctc tccttccagt gtctgggtcc ccgctctttc cttagtgctg 1380

gcagcgggag gggcgcctcc tctgggagac tgccctgacc actgcttttc cttttagtga 1440

gtgatgtgcc atttcctttc tctgcccagt ctggggctgg ggtgccaggc tggggcatcg 1500

cgctgctggt gctggtctgt gttctggttg cgctggccat tgtctatctc attgccttgg 1560

ctgtctgtca gtgccgccga aagaactacg ggcagctgga catctttcca gcccgggata 1620

cctaccatcc tatgagcgag taccccacct accacaccca tgggcgctat gtgcccccta 1680

gcagtaccga tcgtagcccc tatgagaagg tttctgcagg taatggtggc agcagcctct 1740

cttacacaaa cccagcagtg gcagccactt ctgccaactt gtaggggcac gtcgcccgct 1800

gagctgagtg gccagccagt gccattccac tccactcagg ttcttcaggg ccagagcccc 1860

tgcaccctgt ttgggctggt gagctgggag ttcaggtggg ctgctcacag cctccttcag 1920

aggccccacc aatttctcgg acacttctca gtgtgtggaa gctcatgtgg gcccctgagg 1980

gctcatgcct gggaagtgtt gtggtggggg ctcccaggag gactggccca gagagccctg 2040

agatagcggg gatcctgaac tggactgaat aaaacgtggt ctcccactgc 2090

<210> 77
<211> 1808
<212> DNA
<213> Homo sapien

<400> 77
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcagtgagt gatgtgccat ttcctttctc tgcccagtct ggggctgggg 1200

tgccaggctg gggcatcgcg ctgctggtgc tggtctgtgt tctggttgcg ctggccattg 1260

tctatctcat tgccttggct gtctgtcagt gccgccgaaa gaactacggg cagctggaca 1320

tctttccagc ccgggatacc taccatccta tgagcgagta ccccacctac cacacccatg 1380

ggcgctatgt gccccctagc agtaccgatc gtagccccta tgagaaggtt tctgcaggta 1440

atggtggcag cagcctctct tacacaaacc cagcagtggc agccacttct gccaacttgt 1500

aggggcacgt cgcccgctga gctgagtggc cagccagtgc cattccactc cactcaggtt 1560

cttcagggcc agagcccctg caccctgttt gggctggtga gctgggagtt caggtgggct 1620

gctcacagcc tccttcagag gccccaccaa tttctcggac acttctcagt gtgtggaagc 1680

tcatgtgggc ccctgagggc tcatgcctgg gaagtgttgt ggtgggggct cccaggagga 1740

ctggcccaga gagccctgag atagcgggga tcctgaactg gactgaataa aacgtggtct 1800

cccactgc 1808

<210> 78
<211> 1823
<212> DNA
<213> Homo sapien

<400> 78
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt ttttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcaggccag gatctgtggt ggtacaattg actctggcct tccgagaagg 1200

taccatcaat gtccacgacg tggagacaca gttcaatcag tataaaacgg aagcagcctc 1260

tcgatataac ctgacgatct cagacgtcag cggctgtctg tcagtgccgc cgaaagaact 1320

acgggcagct ggacatcttt ccagcccggg atacctacca tcctatgagc gagtacccca 1380

cctaccacac ccatgggcgc tatgtgcccc ctagcagtac cgatcgtagc ccctatgaga 1440

aggtttctgc aggtaatggt ggcagcagcc tctcttacac aaacccagca gtggcagcca 1500

cttctgccaa cttgtagggg cacgtcgccc gctgagctga gtggccagcc agtgccattc 1560

cactccactc aggttcttca gggccagagc cctgcaccc tgtttgggct ggtgagctgg 1620

gagttcaggt gggctgctca cagcctcctt cagaggcccc accaatttct cggacacttc 1680

tcagtgtgtg gaagctcatg tgggcccctg agggctcatg cctgggaagt gttgtggtgg 1740

gggctcccag gaggactggc ccagagagcc ctgagatagc ggggatcctg aactggactg 1800

aataaaacgt ggtctcccac tgc 1823

<210> 79
<211> 1630
<212> DNA
<213> Homo sapien

<400> 79
taggaggtag gggaggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgg ctgtctgtca gtgccgccga aagaactacg ggcagctgga 1140

catctttcca gcccgggata cctaccatcc tatgagcgag taccccacct accacaccca 1200

tgggcgctat gtgcccccta gcagtaccga tcgtagcccc tatgagaagg tttctgcagg 1260

taatggtggc agcagcctct cttacacaaa cccagcagtg gcagccactt ctgccaactt 1320

gtaggggcac gtcgcccgct gagctgagtg gccagccagt gccattccac tccactcagg 1380

ttcttcaggg ccagagcccc tgcaccctgt ttgggctggt gagctgggag ttcaggtggg 1440

ctgctcacag cctccttcag aggccccacc aatttctcgg acacttctca gtgtgtggaa 1500

gctcatgtgg gcccctgagg gctcatgcct gggaagtgtt gtggtggggg ctcccaggag 1560

gactggccca gagagccctg agatagcggg gatcctgaac tggactgaat aaaacgtggt 1620

ctcccactgc 1630

<210> 80
<211> 640
<212> DNA
<213> Homo sapien

<400> 80
agtcgtgacg tggcacaacc ctggcgctgg ggtgccaggc tggggcatcg cgctgctggt 60

gctggtctgt gttctggttg cgctggccat tgtctatctc attgccttgg ctgtctgtca 120

gtgccgccga aagaactacg ggcagctgga catctttcca gcccgggata cctaccatcc 180

tatgagcgag taccccacct accacaccca tgggcgctat gtgcccccta gcagtaccga 240

tcgtagcccc tatgagaagg tttctgcagg taatggtggc agcagcctct cttacacaaa 300

cccagcagtg gcagccactt ctgccaactt gtaggggcac gtcgcccgct gagctgagtg 360

gccagccagt gccattccac tccactcagg ttcttcaggg ccagagcccc tgcaccctgt 420

ttgggctggt gagctgggag ttcaggtggg ctgctcacag cctccttcag aggccccacc 480

aatttctcgg acacttctca gtgtgtggaa gctcatgtgg gcccctgagg gctcatgcct 540

gggaagtgtt gtggtggggg ctcccaggag gactggccca gagagccctg agatagcggg 600

gatcctgaac tggactgaat aaaacgtggt ctcccactgc 640

<210> 81
<211> 874
<212> DNA
<213> Homo sapien

<400> 81
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agttgttaca ggttctggtc atgcaagctc taccccaggt ggagaaaagg agacttcggc 300

tacccagaga agttcagtgc ccagctctac tgagaagaat gctgctgtct gtcagtgccg 360

ccgaaagaac tacgggcagc tggacatctt tccagcccgg gatacctacc atcctatgag 420

cgagtacccc acctaccaca cccatgggcg ctatgtgccc cctagcagta ccgatcgtag 480

cccctatgag aaggtttctg caggtaatgg tggcagcagc ctctcttaca caaacccagc 540

agtggcagcc acttctgcca acttgtaggg gcacgtcgcc cgctgagctg agtggccagc 600

cagtgccatt ccactccact caggttcttc agggccagag cccctgcacc ctgtttgggc 660

tggtgagctg ggagttcagg tgggctgctc acagcctcct tcagaggccc caccaatttc 720

tcggacactt ctcagtgtgt ggaagctcat gtgggcccct gagggctcat gcctgggaag 780

tgttgtggtg ggggctccca ggaggactgg cccagagagc cctgagatag cggggatcct 840

gaactggact gaataaaacg tggtctccca ctgc 874

<210> 82
<211> 1084
<212> DNA
<213> Homo sapien

<400> 82
ttttgctttt ttgcacccag aggcaaaatg ggtggagcac tatgcccagg ggagcccttc 60

ccgaggagtc ccaggggtga gcctctgtgc ccctaatcat ctcctaggaa tggagggtag 120

accgagaaag gctggcatag ggggaggttt cccaggtaga agaagaagtg tcagcagacc 180

aggtgagcgt gggtgccagt ggggttcttg ggagcttcaa ggaagcaagg aacgctccct 240

ccttcctctc ctggtctttc tctatgggac ctagtaaata attactgcag ccacctgagg 300

ctggaaaacc actccaggtg ggggaggaga gagtttagtt ttcttgctcc tattttcctc 360

ctcctggaga cctccctctc tcggctttac aaagacacag atacaccccg ccccccaaac 420

acacacacac acacacacac acacctcctt aggctggaac agcagagaat ggagggacaa 480

gggggctgat tagagccaag aagagggagt gaaggagagc agagggagga gggcagccct 540

gtttacagtc acctggctgg tggggtggca ggtgctctct ctgaattaac cctttgagag 600

ctggccagga ctctggactg attaccccag cctggggtgg catccagggg ctctaggagg 660

taccttttgc tcctcaccct ggatctcttt tccttccacc caggtttctg caggtaatgg 720

tggcagcagc ctctcttaca caaacccagc agtggcagcc acttctgcca acttgtaggg 780

gcacgtcgcc cgctgagctg agtggccagc cagtgccatt ccactccact caggttcttc 840

agggccagag cccctgcacc ctgtttgggc tggtgagctg ggagttcagg tgggctgctc 900

acagcctcct tcagaggccc caccaatttc tcggacactt ctcagtgtgt ggaagctcat 960

gtgggcccct gagggctcat gcctgggaag tgttgtggtg ggggctccca ggaggactgg 1020

cccagagagc cctgagatag cggggatcct gaactggact gaataaaacg tggtctccca 1080

ctgc 1084

<210> 83
<211> 1194
<212> DNA
<213> Homo sapien

<400> 83
taggaggtag gggagggggc ggggttttgt cacctgtcac ctgctccggc tgtgctatgg 60

gcgggcgggc ggggagtggg gggaccggta taaagcggta ggcgcctgtg cccgctccac 120

ctctcaagca gccagcgcct gcctgaatct gttctgcccc ctccccaccc atttcaccac 180

caccatgaca ccgggcaccc agtctccttt cttcctgctg ctgctcctca cagtgcttac 240

agctaccaca gcccctaaac ccgcaacagt tgttacaggt tctggtcatg caagctctac 300

cccaggtgga gaaaaggaga cttcggctac ccagagaagt tcagtgccca gctctactga 360

gaagaatgct gtgagtatga ccagcagcgt actctccagc cacagccccg gttcaggctc 420

ctccaccact cagggacagg atgtcactct ggccccggcc acggaaccag cttcaggttc 480

agctgccacc tggggacagg atgtcacctc ggtcccagtc accaggccag ccctgggctc 540

caccaccccg ccagcccacg atgtcacctc agccccggac aacaagccag ccccgggctc 600

caccgccccc ccagcccacg gtgtcacctc ggccccggac accaggccgg ccccgggctc 660

caccgccccc ccagcccatg gtgtcacctc ggccccggac aacaggcccg ccttgggctc 720

caccgcccct ccagtccaca atgtcacctc ggcctcaggc tctgcatcag gctcagcttc 780

tactctggtg cacaacggca cctctgccag ggctaccaca accccagcca gcaagagcac 840

tccattctca attcccagcc accactctga tactcctacc acccttgcca gccatagcac 900

caagactgat gccagtagca ctcaccatag cacggtacct cctctcacct cctccaatca 960

cagcacttct ccccagttgt ctactggggt ctctttcttt ttcctgtctt ttcacatttc 1020

aaacctccag tttaattcct ctctggaaga tcccagcacc gactactacc aagagctgca 1080

gagagacatt tctgaaatgt tttgcagat ttataaacaa gggggttttc tgggcctctc 1140

caatattaag ttcagccagg agctgtggtg gcagaataag cgatcctcta atca 1194

<210> 84
<211> 2623
<212> DNA
<213> Homo sapien

<400> 84
ctggaatctg gacacacagg gctccccccc gcctctgact tctctgtccg aagtcgggac 60

accctcctac cacctgtaga gaagcgggag tggatctgaa ataaaatcca ggaatctggg 120

ggttcctaga cggagccaga cttcggaacg ggtgtcctgc tactcctgct ggggctcctc 180

caggacaagg gcacacaact ggttccgtta agcccctctc tcgctcagac gccatggagc 240

tggatctgtc tccacctcat cttagcagct ctccggaaga cctttgccca gcccctggga 300

cccctcctgg gactccccgg ccccctgata cccctctgcc tgaggaggta aagaggtccc 360

agcctctcct catcccaacc accggcagga aacttcgaga ggaggagagg cgtgccacct 420

ccctcccctc tatccccaac cccttccctg agctctgcag tcctccctca cagagcccaa 480

ttctcggggg cccctccagt gcaagggggc tgctcccccg cgatgccagc cgcccccatg 540

tagtaaaggt gtacagtgag gatggggcct gcaggtctgt ggaggtggca gcaggtgcca 600

cagctcgcca cgtgtgtgaa atgctggtgc agcgagctca cgccttgagc gacgagacct 660

gggggctggt ggagtgccac ccccacctag cactggagcg ggtttggag gaccacgagt 720

ccgtggtgga agtgcaggct gcctggcccg tgggcggaga tagccgcttc gtcttccgga 780

aaaacttcgc caagtacgaa ctgttcaaga gctccccaca ctccctgttc ccagaaaaaa 840

tggtctccag ctgtctcgat gcacacactg gtatatccca tgaagacctc atccagaact 900

tcctgaatgc tggcagcttt cctgagatcc agggctttct gcagctgcgg ggttcaggac 960

ggaagctttg gaaacgcttt ttctgcttct tgcgccgatc tggcctctat tactccacca 1020

agggcacctc taaggatccg aggcacctgc agtacgtggc agatgtgaac gagtccaacg 1080

tgtacgtggt gacgcagggc cgcaagctct acgggatgcc cactgacttc ggtttctgtg 1140

tcaagcccaa caagcttcga aatggccaca aggggcttcg gatcttctgc agtgaagatg 1200

agcagagccg cacctgctgg ctggctgcct tccgcctctt caagtacggg gtgcagctgt 1260

acaagaatta ccagcaggca cagtctcgcc atctgcatcc atcttgtttg ggctccccac 1320

ccttgagaag tgcctcagat aataccctgg tggccatgga cttctctggc catgctgggc 1380

gtgtcattga gaacccccgg gaggctctga gtgtggccct ggaggaggcc caggcctgga 1440

ggaagaagac aaaccaccgc ctcagcctgc ccatgccagc ctccggcacg agcctcagtg 1500

cagccatcca ccgcacccaa ctctggttcc acgggcgcat ttcccgtgag gagagccagc 1560

ggcttattgg acagcagggc ttggtagacg gcctgttcct ggtccgggag agtcagcgga 1620

acccccaggg ctttgtcctc tctttgtgcc acctgcagaa agtgaagcat tatctcatcc 1680

tgccgagcga ggaggagggc cgcctgtact tcagcatgga tgatggccag acccgcttca 1740

ctgacctgct gcagctcgtg gagttccacc agctgaaccg cggcatcctg ccgtgcttgc 1800

tgcgccattg ctgcacgcgg gtggccctct gaccaggccg tggactggct catgcctcag 1860

cccgccttca ggctgcccgc cgcccctcca cccatccagt ggactctggg gcgcggccac 1920

aggggacggg atgaggagcg ggagggttcc gccactccag ttttctcctc tgcttctttg 1980

cctccctcag atagaaaaca gcccccactc cagtccactc ctgacccctc tcctcaaggg 2040

aaggccttgg gtggccccct ctccttctcc tagctctgga ggtgctgctc tagggcaggg 2100

aattatggga gaagtggggg cagcccaggc ggtttcacgc cccacacttt gtacagaccg 2160

agaggccagt tgatctgctc tgttttatac tagtgacaat aaagattatt ttttgataaa 2220

aaactcagaa ctatctcgtc gcgagtttga taaaaagtgt aaaaaaactg gggggaactt 2280

catagggggt caaacatctc gctgccggcg gataggactt ggctaaactt cttccgagcg 2340

ggccccgtaa gggtggtatg ctgataaaaa tggggggggg cccccctctc agggggccct 2400

ccagaacctt ttgggggtgg ggtacccttg ggtggttaac tagtgaactc tttcctcaaa 2460

aggttgccgc cccctgtgta ttgtcgacaa ttttcttggg gggcgggccc gttttctttt 2520

caccacgctt ttgttttccc gggtggggaa cccacccctg gtgtgtgtgc cccccccgtt 2580

tattttgggc gccctttttg tgggggggaaa ttcccccgct ttt 2623

<210> 85
<211> 1036
<212> DNA
<213> Homo sapien

<400> 85

ctgagaggca gcgaactcat ctttgccagt acaggagctt gtgccgtggg cccacagccc 60

acagcccaca gccatggtaa ggcagatgtc acaggtgggg ggaggtgggc tctgtgccag 120

ccaattttcg tctccctccc ccagccaagg tctcccaggg gtgcagggag agcggagctg 180

ctcagagctt ggccaggttc taagtgtgct cctgaaagca ggtcacccct gagatcctca 240

gggtggggca cagaggggca ccctagcagg taaagggagg ccacgggatg gcggtgggca 300

gctggccttc tagtaacgag ccctcagtgc cttctgtgcc tggggtccct gccgacggga 360

tgtagaggac agacaggagg gagcactgtc cctgggtaca ggagctcgcc ctgcagccag 420

tgccttgtgt gtggtgggcc tggggctggc gccgcagtct ctgaacctgt gtgacgcctg 480

cagggctggg acctgacggt gaagatgctg gcgggcaacg aattccaggt gtccctgagc 540

agctccatgt cggtgtcaga gctgaaggcg cagatcaccc agaagatcgg cgtgcacgcc 600

ttccagcagc gtctggctgt ccacccgagc ggtgtggcgc tgcaggacag ggtcccccctt 660

gccagccagg gcctgggccc cggcagcacg gtcctgctgg tggtggacaa atgcgacgaa 720

cctctgagca tcctggtgag gaataacaag ggccgcagca gcacctacga ggtgcggctg 780

acgcagaccg tggcccacct gaagcagcaa gtgagcgggc tggagggtgt gcaggacgac 840

ctgttctggc tgaccttcga ggggaagccc ctggaggacc agctcccgct gggggagtac 900

ggcctcaagc ccctgagcac cgtgttcatg aatctgcgcc tgcggggagg cggcacagag 960

cctggcgggc ggagctaagg gccccaccag catccgagca ggatcaaggg ccggaaataa 1020

aggctgttgt aaagag 1036

<210> 86
<211> 753
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (168)..(208)
<223> n=a, c, g, or t

<400> 86
gctgcctcta taggtgctgg tatataagta ttatcgacat catttaagta atgatttaga 60

agttacataa aaaaaaaatt tccccaagtt attttctggc gaagagcttc cctggtatga 120

cctgaaactc aaacttggaa aagagataaa tttaattgga taaaaatnnn nnnnnnnnnn 180

nnnnnnnnnn nnnnnnnnnn nnnnnnnntc tcctgaatct tttatctatg ccttaagcct 240

tttctgttcc cttcaggacc taggcttttg aaacccaaaa gccaggaaaa catgcctttg 300

ttatctgctt tctgcaatca cgtctcttcc atggggcact gagcagagaa tggtgtggcc 360

aagtgagtag tgagaagcag tgaggaggtg tgagctaggt gtctgttccc attttagaaa 420

atactgttcc tacatcagaa ataccacatt aagacgtata gagccaggtc actgggatgc 480

ttgaacccaa atagctggga ttctggacag agtcagcaga gtacagaagg ctctgaagtg 540

ggagacggag ctggggtgca tccctcccag tgaggagggg tcatgagggg cgtctgggaa 600

gagggacatt tgaactagga ttagctgagt tgccatgatg ctaagataat gggagagtgt 660

tctttgtggt caccagtgtc cacatggcat cccttccctg agattttcat cactccctgt 720

ggtcttcagt cagtaaagct cttagaacac ttg 753

<210> 87
<211> 878
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (282)..(322)
<223> n=a, c, g, or t

<400> 87
cggaggccga ggttgcggtg agctaagatc gtacccttgc actccagcct gggtgacgga 60

gtaagactcc atctccaaaa agaaaagaag aattgatatt gatattggaa gggagctgcc 120

tctataggtg ctggtatata agtattatcg acatcattta agtaatgatt tagaagttac 180

ataaaaaaaa aatttcccca agttattttc tggcgaagag cttccctggt atgacctgaa 240

actcaaactt ggaaaagaga taaatttaat tggataaaaa tnnnnnnnnn nnnnnnnnnn 300

nnnnnnnnnn nnnnnnnnnn nntctcctga atcttttatc tatgccttaa gccttttctg 360

ttcccttcag gacctaggct tttgaaaccc aaaagccagg aaaacatgcc tttgttatct 420

gctttctgca atcacgtctc ttccatgggg cactgagcag agaatggtgt ggccaagtga 480

gtagtgagaa gcagtgagga ggtgtgagct aggtgtctgt tcccatttta gaaaatactg 540

ttcctacatc agaaatacca cattaagacg tatagagcca ggtcactggg atgcttgaac 600

ccaaatagct gggattctgg acagagtcag cagagtacag aaggctctga agtgggagac 660

ggagctgggg tgcatccctc ccagtgagga ggggtcatga ggggcgtctg ggaagaggga 720

catttgaact aggattagct gagttgccat gatgctaaga taatgggaga gtgttctttg 780

tggtcaccag tgtccacatg gcatcccttc cctgagattt tcatcactcc ctgtggtctt 840

cagtcagtaa agctcttaga acacttaaaa aaaaaaa 878

<210> 88
<211> 1020
<212> DNA

<213> Homo sapien

<400> 88
caaatgcaca gtccccctcc cactccgtta cctaactgta cgtcttttca tgtttataaa 60

ctatacagaa aactgtattt gctgaactaa ggattgtatt ggtgatttct agcaaaaaca 120

aagtgataga atttttgtct agaatcccaa actggcaacg atagtctcca agggacctgg 180

ccttgccaag ggcctggggc aaggtgtcgg cgggacggtg aggaaggggg aggcagcaag 240

agtcactttg gggggaccaat attcttagat atttagagca tcaccttgtt tttatatgca 300

acacaagcct gtctgccacc ctggagcgcc ctgtcacccc tgctgtcgta gctgttggct 360

tcagggtgag aagtgagaag cagcttattg tatatgaggg agccaggccc cgagggtgag 420

cgagatggag aaggggaagg aaggggcttt gggatctgga aaccagcagg ccaggcagca 480

tccacagtgt tagtccaaag ggtcggaccg tgtcgtcagc ctagcgtttg gtcagtgacg 540

gcctggacgg gccaaggaga ctccgggctt gagcccaggc ctcccgcacg gctcagctgc 600

tgaatttttc cttgaggctg tttggtgtgt gacccagcaa gggccctgtg tgggacagca 660

ggagggaggc gtcgcggggc cttagcagaa ggggaacaat gagggcattt catgaaccat 720

ctcaggcact tctgcatcac ggaagacctg gccctcccag ccgtcctggg gatgctcagg 780

gtgcaggcag aggctcggga ggccggactc aggggtcaga agcagggact ggggcaggcg 840

agcccggaca gggaagaggg gctccgatca aagccggccg tgctgctggc cgggggccca 900

ggtgggtaca agctcctttg tgctttgcac aaacctgaat cccccaccag agaggatgtg 960

tgtgaggagc cagaaacgct gaatccaatt aagagagaaa aataataata acgaatgacg 1020

<210> 89
<211> 1854
<212> DNA
<213> Homo sapien

<400> 89
ctggggctgg cggtcactct ccgctgagga cccagggcgt cacacccagc actgccacat 60

gtccaccaag gaacagaatt tattttcttc ttttttaac aagtggaaga tctgctgggt 120

ttcaggaaaa ggctggtaga ggcttcggct gctgtctgga cgtctggacc ctgccatgtg 180

gattataaac ccaaagtgta cagccctagg cgggaggggg tggcgcttct cagccggctg 240

tcccagccag ccccgcagag cgcccacgga cagtgtccac tctggcaagg tgggaaaagg 300

cactccaagt gcatcctcca ctggcaacag tgggacaatt gcccccgacg gcggcaccgg 360

ggctctgtgg aatcccgatc gttccgagag gtctggaggg ccccgtggtt cctggagaaa 420

gcaggacgca gagaagaaca aatgaggctc acccacgagg ctgggtggcc agcagtctgg 480

gcacacacga gcaggtggca tcttggctct tgcctgaggc cagtcaccct gccctgaatt 540

ctaccctact ccaccttcag cccctcccgc gggggtagcg cctctcattc ctgatgtctc 600

aggcaaccct ggcagaccca ggtccaactg ctggggtcca agaaccaatt accaaaggaa 660

agatcatcag aggctgaaat ctagaacttc atcccgggca atgaggttct cacagaaggt 720

gcagttttat aactaactac gtccacttat atatattcac actctacata tatatatata 780

tatatatata tatatatata tatatataca cacaaatgca cagtcccccct cccactccgt 840

tacctaactg tacgtctttt catgtttata aactatacag aaaactgtat ttgctgaact 900

aaggattgta ttggtgattt ctagcaaaaa caaagtgata gaatttttgt ctagaatccc 960

aaactggcaa cgatagtctc caagggacct ggccttgcca agggcctggg gcaaggtgtc 1020

ggcgggacgg tgaggaaggg ggaggcagca agagtcactt tgggggacca atattcttag 1080

atatttagag catcaccttg tttttatatg caacacaagc ctgtctgcca ccctggagcg 1140

ccctgtcacc cctgctgtcg tagctgttgg cttcagggtg agaagtgaga agcagcttat 1200

tgtatatgag ggagccaggc cccgagggtg agcgagatgg agaaggggaa ggaaggggct 1260

ttgggatctg gaaaccagca ggccaggcag catccacagt gttagtccaa agggtcggac 1320

cgtgtcgtca gcctagcgtt tggtcagtga cggcctggac gggccaagga gactccgggc 1380

ttgagcccag gcctcccgca cggctcagct gctgaatttt tccttgaggc tgtttggtgt 1440

gtgacccagc aagggccctg tgtgggacag caggagggag gcgtcgcggg gccttagcag 1500

aaggggaaca atgagggcat ttcatgaacc atctcaggca cttctgcatc acggaagacc 1560

tggccctccc agccgtcctg ggatgctca gggtgcaggc agaggctcgg gaggccggac 1620

tcaggggtca gaagcaggga ctggggcagg cgagcccgga cagggaagag gggctccgat 1680

caaagccggc cgtgctgctg gccgggggcc caggtgggta caagctcctt tgtgctttgc 1740

acaaacctga atcccccacc agagaggatg tgtgtgagga gccagaaacg ctgaatccaa 1800

ttaagagaga aaaataataa taacaataaa tgatcttgga caagaaaaaa aaaa 1854

<210> 90
<211> 1759
<212> DNA
<213> Homo sapien

<400> 90
atgtgaaaag aaaatagtta tctgtgcttg gtgttgtgtg ctctcctaaa gttaaccaga 60

cgtgaagcca aaaacatcaa ctgggactga caacacaaga aagattcttt aactgaggtg 120

gttaaatggc cctgaaaaga gcctttggag acaaagcagc cggcgacccg cggagggagg 180

gagggaggga gcgagcgagc gccaggtccc ggcagggact cacttggagc tggcgtactt 240

ggtgaccgcc ttggtgccct cggacacggc gtgcttggcc agctcgccgg gcagcagcag 300

gcgcacggcc gtctggatct ccgggacgtg atggtggagc gcttgttgta gtgcgccagg 360

cgggacgcct ctcccgcgat gcgctcgaag atgtcgttga ggaaggagtt catgatgccc 420

atggccttgc accagatgcc ggtgtcgggg tggacccgct tcagcacctt gtacacgtag 480

atggagtagc tctccttgcg gctgcgcttg cgcttcttgc cgtctttctt ctgggctttg 540

gtgacggctt tcttggagcc cttcttggga gccggcgcga actttgcagg ctcaggcatg 600

gccagaccca agaccgacac cgacccccga gaacgcaagc agagcggtag gctcggggtc 660

taccggaaac gactgtgtac ttacagaggc tgtgcgcatg acgctgcgtt atggttcgcg 720

agttttccgc ggcgcgcaat gcgagggaga cgagattatg taaatgagtg gattctggct 780

gagctatcct attggctatc gggacaaaat ttgcttgagc caatcaaagt gctccgtgga 840

caatcgccgt tctgtctata aaaaggtgaa gcagcggcgt tttcggcgac tttcccgatc 900

gccaggcagg agtttctctc ggtgactact atcgctgtca tgtctggtcg tggcaagcaa 960

ggaggcaagg cccgcgccaa ggccaagtcg cgctcgtccc gcgctggcct tcagttcccg 1020

gtagggcgag tgcatcgctt gctgcgcaaa ggcaactacg cggagcgagt gggggccggc 1080

gcgcccgtct acatggctgc ggtcctcgag tatctgaccg ccgagatcct ggagctggcg 1140

ggcaacgcgg ctcgggacaa caagaagacg cgcatcatcc ctcgtcacct ccagctggcc 1200

atccgcaacg acgaggaact gaacaagctg ctgggcaaag tcaccatcgc ccagggcggc 1260

gtcttgccta acatccaggc cgtactgctc cctaagaaga cggagagtca ccacaaggca 1320

aagggcaagt gaggctgacg tccggcccaa gtgggcccag cccggcccgc gtctcgaagg 1380

ggcacctgtg aactcaaaag gctcttttca gagccaccca cgttttcaaa taaaagagtt 1440

gttaatgctg gccactctca gtccagcgtt cctcagtagt gaatagcgaa cctggagctg 1500

acgggacggg acgggacggg acgggacggg gcggggcggg gcggggcggg gtgtgtgtgt 1560

gtgcgcgccg tcttccatct ggagcacgta actgccttgg ctcttcgatg agtgggtccc 1620

cagtcctagg acttcccagg gcaggtgcag gcaccaaacg tcctgggcgc cgccacggtc 1680

cgctccacac agtcacaaac accagcgccg cgggcagtac ccaacgcgct gaagtgttgc 1740

gcgcggagcg cgcgcttcc 1759

<210> 91
<211> 1234
<212> DNA
<213> Homo sapien

<400> 91
ggtcactctc tactcaagtt ctacttatat aacagcaatg cagctctctt cataaagctg 60

gctgttgtgt agtttatgtt ggggaatcag ttcatggttt aaaaagttct gtcaatgcag 120

agaacaagcc ggtgtgtttt atggagaggc tgtttaatct ccactgtgag acagtaaata 180

tttggctgtt gcatcatcgt gaagcttatg atcacagtct ggcgccatct ccctcctcgc 240

ctggagtctg atctgtcccg gcccagtgtc ctccaggaac ctggcccctc atgcctccgt 300

gcttgcgcgt gtgccatttc ctctctccag aggacctttc ctgcctagga ctcatcattg 360

tccccttcct ggtaagccat ccccgacctt ccaggcagaa cctgctggct tctcctcagc 420

actttgcatg gatttcatgt cacagtcctg ggtgcactgt gtcgccctct ctatgtgtca 480

gcctcccgtc ccctaccgtg ggctcctcca gggaggtgtg gacattcatc ctcttccagg 540

cagccctcag gaatccaggg agaagataag gaggcggggc gggcggaggg gggtgctcca 600

cacactcaga acactttcct ctgcacttac ttcattctgg tttttctttt gggtccttgg 660

tgtttttaaa taaacccttt cctgtagttt gctccccttc catggagggc tgtttcgagc 720

acagatctgc tgggtgtctg tatttacaaa gagaaggggc cactcgtgtg tgagcagcac 780

cgagggacag aggtaccttg cctgcttgtg tcccctccaa gtccttctga tattttcctt 840

tccagctgtt gcctagtttc ctggtattaa ggagaatcaa ctctctggat aaacgtggta 900

aatatggccc atagtcccat ctttttacag gcattttta cacctggagc agccagagga 960

cgcatgcatg gctcttcgga aggtaattta gggatcaccc atgtaagttt cctaaggatt 1020

tctttaacat ggttcttctg attcagtccg gccaattaaa tctaaatcca cccctgaaag 1080

ccatctggtg tggataacaa gcccacaaat gagcagtcag ctttttgtgc cctttagggc 1140

ctgggacaac cacgggatct aaaaggggct ggaactagag gtcttgagct cctgttccta 1200

aaatcatctt catcctatat ctgcagtctt ctcc 1234

<210> 92
<211> 730
<212> DNA
<213> Homo sapien

<400> 92
cagcgtcaga gagaaagaac tgactgaaac gtttgagata tataggaaac atcaaaaggt 60

gataaaattt ccctagaatc tccactatct caaagatgaa gaaagttctc ctcctgatca 120

cagccatctt gggcagtggc tgttggtttc ccagtctctc aagaccagga acgagaaaaa 180

agaagtgtaa gttacctttt ctctttttta catatcagtg acagcgatga attagcttca 240

gggttttttg tgttccctta cccatatcca tttcgcccac ttccaccaat tccatttcca 300

agatttccat ggtttagacg taattttcct attccaatac ctgaatctgc ccctacaact 360

ccccttccta gcgaaaagta aacaagaagg aaaagtcacg ataaacctgg tcacctgaaa 420

ttgaaattga gccacttcct tgaagaatca aaattcctgt taataaaaga aaaacaaatg 480

taattgaaat agcacacagc attctctagt caatatcttt agtgatcttc tttaataaac 540

atgaaagcaa atcactaaag atattgacta gagaatgctg tgtgctattt caatatcttt 600

agtgatcttc tttaataaac atgaaagcat aaaaaaaaaa agacgaaaaa aaaaggctgg 660

gggcaccctg ggacaaagcg gtcccggggg ggattggttc ccggccaatt ccacaataag 720

ccgcacaaga 730

<210> 93
<211> 1159
<212> DNA
<213> Homo sapien

<400> 93
ggggacagat ttctccattc cattatacct ttgagtatat aaaacagcta caatattcca 60

gggccagtca cttgccattt ctcataacag cgtcagagag aaagaactga ctgaaacgtt 120

tgagatatat aggaaacatc aaaaggtgat aaaatttccc tagaatctcc actatctcaa 180

agatgaagaa agttctcctc ctgatcacag ccatcttggc agtggctgtt ggtttcccag 240

tctctcaaga ccaggaacga gaaaaaagaa gtgtaagtta ccttttctct tttttacata 300

tcagtgacag cgatgaatta gcttcagggt tttttgtgtt cccttaccca tatccatttc 360

gcccacttcc accaattcca tttccaagat ttccatggtt tagacgtaat tttcctattc 420

caatacctga atctgcccct acaactcccc ttcctagcga aaagtaaaca agaaggaaaa 480

gtcacgataa acctggtcac ctgaaattga aattgagcca cttccttgaa gaatcaaaat 540

tcctgttaat aaaagaaaaa caaatgtaat tgaaatagca cacagcattc tctagtcaat 600

atctttagtg atcttcttta ataaacatga aagaagatca ctaaagatat tgactagaga 660

atgctgtgtg ctatttcaat tacatttgtt tttctttaat aaacatgaat tttgattctt 720

caaggaagtg gctcaatttc aatttcaggt gacctgaaat aaataacaga catatggtta 780

ttaattgcaa tgggtcattt tcttggaaac atatacattt tctgcatttt aatgacaact 840

attggcttaa aaatatatct agttcaagga ctgggaaacc atctgctcaa gatgtagaaa 900

gaaagcaaag gtctttagtg gtaagtagta gctgaaatat ttttttccta gaacagtcct 960

ctgggttcta atttaatctt agataagatt aaattatata tattaaatta taaattatta 1020

tagtagatta gatctatagt ctatagtata gattatattt cctcaattta tctagtaatt 1080

gacacaccat ccactttgtt tttgatgtga tgaaatgaca ggggccactg ttataggtga 1140

agcatgaagc ctttaaaat 1159

<210> 94
<211> 1493

<212> DNA
<213> Homo sapien

<400> 94
ggagcccagc cgtgggattt tcaggtgttt tcatttggtg atcaggactg aacagagaga 60

actcaccatg gagtttgggc tgagctggct ttttcttgtg gctattttaa aaggtgtcca 120

gtgtgaggtg cagctgttgg agtctggggg aggcttggta cagcctgggg ggtccctgag 180

actctcctgt gcagcctctg gattcacctt tagcatctat gccatgagct gggtccgcca 240

ggctccaggg aaggggctgg agtgggtcgc aagtatcagt ttcagtggtg gtagtacata 300

ctacgcagac tccgtgaagg gccgtttcac catctccaga gacaattcca agaccacgat 360

gcatctccac atgaacagcc tgagaaccga cgacacggcc gtctactact gtgcgaaacc 420

gtttccgtat tttgactact ggggccaggg aaccctggtc accgtctcga gtggcgatgg 480

gtccagtggc ggtagcgggg gcgcgtcgac tggcgaagtt gtgttgacgc agtttccagg 540

gcaccctgtc tctgtctcca ggggaaagag ccaccctctc ctgcagggcc agtcagagtg 600

cttagcagca gctacttagc ctggtatcag cagagacctg gccaggctcc caggctcctc 660

gtttatagtg catctgtgcg gcccaatgat attccagtca gggtccgtgg cagtgggtct 720

gggacagagt tcactctcac catcagcaga ctggtaacct gaagattttg cagtgtatta 780

ctgtcaacag ctatgggggc tcacctgacg tggactttcg ccccggggac caaggtggaa 840

gtccaaacga actgtggctg caccatctgt cttcatcttc ccgccatctg atgagcagtt 900

gaaatctgga actgcctctg ttgtgtgcct gctgaataac ttctatccca gagaggccaa 960

agtacagtgg aaggtggata acgccctacc aatcgggtaa ctcccaggag agtgtcacag 1020

agcaggacag cacaggacag acacctacag cctcagcagc accctgacgc tgagcaaagc 1080

agactacgag aaacacaaac tctacgcctg cgaagtcacc catcagggcc tgagctcgcc 1140

cgtcacaaag agcttcaaca ggggagagtg ttagagggag aagtgccccc acctgctcct 1200

cagttccagc ctgacccccct cccatccttt ggcctctgac cctttttcca caggggacct 1260

acccctattg cggtcctcca gctcatcttt cacctcaccc ccctcctcct ccttggcttt 1320

aattatgcta atgttggagg gagcctgact aaataaagtg aatctttaaa acacaaaaaa 1380

aaggaaaaca aaaaaacaaa aaaaaaaaaa acacgcgggc ggacacccgg ggacaacggg 1440

gtccccgggg tcacactggt tacccgtcca atttcccaca aaacacccgg acc 1493

<210> 95
<211> 177
<212> PRT
<213> Homo sapien

<400> 95

Met Asn Ser Gly Lys Arg Arg Leu Pro Trp Arg Leu Arg Ser Gly Val
1 5 10 15

Pro Ser Pro Pro Gly Leu Leu Ala Pro Ala Pro Ala Pro Cys Ala Pro
20 25 30

Gly Gly His Arg Arg Ala Pro Gly Pro Arg Arg Val Arg Glu Thr Pro
35 40 45

Arg Thr Gly Gly Gly Ile Gly Pro Pro Ser Phe Gly Gly Gly Lys Gly
50 55 60

Gly Trp Lys Glu Glu Gly Ser Gly Val Gly Glu Ser Trp Ser Phe Gly
65 70 75 80

Ile Phe Ser Pro Gly Gln Ala Val Leu Arg Ala Leu Arg Cys Val Ser
85 90 95

Lys Cys Trp Glu Asp Ser Ala Gly Lys Gly Leu Arg Thr Arg Pro Ala
100 105 110

Gly Thr Gly Val Ala Ala Ser Glu Gly Arg Gly Glu Pro Met Ala Ser
115 120 125

Arg Leu Trp Thr Arg Arg Pro Ser Pro Gly Arg Ser Ala Arg Ser Pro
130 135 140

Pro Pro Ala Ser Cys Ala Gly Pro Cys Pro Ala Ser Pro Ala Met Val
145 150 155 160

Pro His Pro Pro Pro Arg Glu Arg Pro Cys Pro Pro Ile Leu His Phe
165 170 175

Pro

<210> 96
<211> 55
<212> PRT
<213> Homo sapien

<400> 96

Met Gln Asn Ser Thr Ser Ser Gly Leu Cys Val Asn Val Pro Pro Phe
1 5 10 15

Pro Pro Leu Ser Gly Cys Leu Asn Val Phe Pro Phe Phe His Leu Lys

20 25 30

Leu Cys Leu Asp Val Leu His Cys His His Leu Phe Leu Arg Lys Arg
35 40 45

Cys Val Pro His Pro Asn Pro
50 55

<210> 97
<211> 24
<212> PRT
<213> Homo sapien

<400> 97

Met Asp His Phe Tyr Leu Leu Ser Asp Thr Tyr Leu Leu Gly Cys Glu
1 5 10 15

Pro Gln Gly Gly Leu Leu Leu Gly
20

<210> 98
<211> 646
<212> PRT
<213> Homo sapien

<400> 98

Met Glu Pro Ala Ala Gly Phe Leu Ser Pro Arg Pro Phe Gln Arg Ala
1 5 10 15

Ala Ala Ala Pro Ala Pro Pro Ala Gly Pro Gly Pro Pro Pro Ser Ala
20 25 30

Leu Arg Gly Pro Glu Leu Glu Met Leu Ala Gly Leu Pro Thr Ser Asp
35 40 45

Pro Gly Arg Leu Ile Thr Asp Pro Arg Ser Gly Arg Thr Tyr Leu Lys
50 55 60

Gly Arg Leu Leu Gly Lys Gly Gly Phe Ala Arg Cys Tyr Glu Ala Thr
65 70 75 80

Asp Thr Glu Thr Gly Ser Ala Tyr Ala Val Lys Val Ile Pro Gln Ser
85 90 95

Arg Val Ala Lys Pro His Gln Arg Glu Lys Ile Leu Asn Glu Ile Glu
100 105 110

Leu His Arg Asp Leu Gln His Arg His Ile Val Arg Phe Ser His His
115 120 125

Phe Glu Asp Ala Asp Asn Ile Tyr Ile Phe Leu Glu Leu Cys Ser Arg
130 135 140

Lys Ser Leu Ala His Ile Trp Lys Ala Arg His Thr Leu Leu Glu Pro
145 150 155 160

Glu Val Arg Tyr Tyr Leu Arg Gln Ile Leu Ser Gly Leu Lys Tyr Leu
165 170 175

His Gln Arg Gly Ile Leu His Arg Asp Leu Lys Leu Gly Asn Phe Phe
180 185 190

Ile Thr Glu Asn Met Glu Leu Lys Val Gly Asp Phe Gly Leu Ala Ala
195 200 205

Arg Leu Glu Pro Pro Glu Gln Arg Lys Lys Thr Ile Cys Gly Thr Pro
210 215 220

Asn Tyr Val Ala Pro Glu Val Leu Leu Arg Gln Gly His Gly Pro Glu
225 230 235 240

Ala Asp Val Trp Ser Leu Gly Cys Val Met Tyr Thr Leu Leu Cys Gly
245 250 255

Ser Pro Pro Phe Glu Thr Ala Asp Leu Lys Glu Thr Tyr Arg Cys Ile
260 265 270

Lys Gln Val His Tyr Thr Leu Pro Ala Ser Leu Ser Leu Pro Ala Arg
275 280 285

Gln Leu Leu Ala Ala Ile Leu Arg Ala Ser Pro Arg Asp Arg Pro Ser
290 295 300

Ile Asp Gln Ile Leu Arg His Asp Phe Phe Thr Lys Gly Tyr Thr Pro
305 310 315 320

Asp Arg Leu Pro Ile Ser Ser Cys Val Thr Val Pro Asp Leu Thr Pro
325 330 335

Pro Asn Pro Ala Arg Ser Leu Phe Ala Lys Val Thr Lys Ser Leu Phe
340 345 350

Gly Arg Lys Lys Lys Ser Lys Asn His Ala Gln Glu Arg Asp Glu Val

355 360 365

Ser Gly Leu Val Ser Gly Leu Met Arg Thr Ser Val Gly His Gln Asp
370 375 380

Ala Arg Pro Glu Ala Pro Ala Ala Ser Gly Pro Ala Pro Val Ser Leu
385 390 395 400

Val Glu Thr Ala Pro Glu Asp Ser Ser Pro Arg Gly Thr Leu Ala Ser
405 410 415

Ser Gly Asp Gly Phe Glu Glu Gly Leu Thr Val Ala Thr Val Val Glu
420 425 430

Ser Ala Leu Cys Ala Leu Arg Asn Cys Ile Ala Phe Met Pro Pro Ala
435 440 445

Glu Gln Asn Pro Ala Pro Leu Ala Gln Pro Glu Pro Leu Val Trp Val
450 455 460

Ser Lys Trp Val Asp Tyr Ser Asn Lys Phe Gly Phe Gly Tyr Gln Leu
465 470 475 480

Ser Ser Arg Arg Val Ala Val Leu Phe Asn Asp Gly Thr His Met Ala
485 490 495

Leu Ser Ala Asn Arg Lys Thr Val His Tyr Asn Pro Thr Ser Thr Lys
500 505 510

His Phe Ser Phe Ser Val Gly Ala Val Pro Arg Ala Leu Gln Pro Gln
515 520 525

Leu Gly Ile Leu Arg Tyr Phe Ala Ser Tyr Met Glu Gln His Leu Met
530 535 540

Lys Gly Gly Asp Leu Pro Ser Val Glu Glu Val Glu Val Pro Ala Pro
545 550 555 560

Pro Leu Leu Leu Gln Trp Val Lys Thr Asp Gln Ala Leu Leu Met Leu
565 570 575

Phe Ser Asp Gly Thr Val Gln Val Asn Phe Tyr Gly Asp His Thr Lys
580 585 590

Leu Ile Leu Ser Gly Trp Glu Pro Leu Leu Val Thr Phe Val Ala Arg
595 600 605

Asn Arg Ser Ala Cys Thr Tyr Leu Ala Ser His Leu Arg Gln Leu Gly
610 615 620

Cys Ser Pro Asp Leu Arg Gln Arg Leu Arg Tyr Ala Leu Arg Leu Leu
625 630 635 640

Arg Asp Arg Ser Pro Ala
645

<210> 99
<211> 99
<212> PRT
<213> Homo sapien

<400> 99

Met Leu Thr Ser Pro Ser Thr Tyr Val Ile Gln Glu Asn Gly Ser Leu
1 5 10 15

Val Glu Ile Arg Asn Ile Leu Gly Glu Lys Tyr Ile Arg Arg Val Arg
20 25 30

Met Arg Pro Gly Val Ala Cys Ser Val Ser Gln Ala Gln Lys Asp Glu
35 40 45

Leu Ile Leu Glu Gly Asn Asp Ile Glu Leu Val Ser Asn Ser Ala Cys
50 55 60

Phe Gly Cys Gln Gln Met Pro Gln Ser Val Lys Asn Lys Asp Ile Arg
65 70 75 80

Lys Phe Leu Asp Gly Ile Tyr Val Ser Glu Lys Gly Thr Val Gln Gln
85 90 95

Ala Asp Glu

<210> 100
<211> 220
<212> PRT
<213> Homo sapien

<400> 100

Met Lys Thr Ile Leu Ser Asn Gln Thr Val Asp Ile Pro Glu Asn Gly
1 5 10 15

Met Arg Leu Asp Val Phe Tyr Leu His Leu Tyr Cys Thr Phe Gln Ala

20 25 30

Leu Cys Gly Leu Thr Ser Val Phe Ser Leu Leu Val Asp Ile Thr Leu
35 40 45

Lys Gly Arg Thr Val Ile Val Lys Gly Pro Arg Gly Thr Leu Arg Arg
50 55 60

Asp Phe Asn His Ile Asn Val Glu Leu Ser Leu Leu Gly Lys Lys Lys
65 70 75 80

Lys Arg Leu Arg Val Asp Lys Trp Trp Gly Asn Arg Lys Glu Leu Ala
85 90 95

Thr Val Arg Thr Ile Cys Ser His Val Gln Asn Met Ile Lys Gly Val
100 105 110

Thr Leu Gly Phe Arg Tyr Lys Met Arg Ser Val Tyr Ala His Phe Pro
115 120 125

Ile Asn Val Val Ile Gln Glu Asn Gly Ser Leu Val Glu Ile Arg Asn
130 135 140

Phe Leu Gly Glu Lys Tyr Ile Arg Arg Val Arg Met Arg Pro Gly Val
145 150 155 160

Ala Cys Ser Val Ser Gln Ala Gln Lys Asp Glu Leu Ile Leu Glu Gly
165 170 175

Asn Asp Ile Glu Leu Val Ser Asn Ser Ala Ala Leu Ile Gln Gln Ala
180 185 190

Thr Thr Val Lys Asn Lys Asp Ile Arg Lys Phe Leu Asp Gly Ile Tyr
195 200 205

Val Ser Glu Lys Gly Thr Val Gln Gln Ala Asp Glu
210 215 220

<210> 101
<211> 47
<212> PRT
<213> Homo sapien

<400> 101

Met Arg Trp His Thr Tyr Leu Cys Cys Leu Lys Val Thr Ile Met Leu
1 5 10 15

Pro Tyr Gln Ala Glu Asn Val Thr Thr Ile Trp Arg Phe Arg Arg Val
20 25 30

Phe Leu Ser Glu Ser Val Met Asn Thr Leu Val Gly Trp Ile Gln
35 40 45

<210> 102
<211> 51
<212> PRT
<213> Homo sapien

<400> 102

Met Ser Ser His Lys Thr Phe Arg Ile Lys Arg Phe Leu Ala Lys Lys
1 5 10 15

Gln Lys Gln Asn Arg Pro Ile Pro Gln Trp Ile Arg Met Lys Thr Gly
20 25 30

Asn Lys Ile Arg Tyr Asn Ser Lys Arg Arg His Trp Arg Arg Thr Lys
35 40 45

Leu Gly Leu
50

<210> 103
<211> 53
<212> PRT
<213> Homo sapien

<400> 103

Met Glu Arg Val Leu Glu Lys Gln Glu Lys Lys Ser Cys Leu Lys Pro
1 5 10 15

His Val Tyr Cys Arg His Arg Arg Glu Trp Arg His Leu Ser Ile Leu
20 25 30

Phe Ser Ile Ser Thr Ala Pro Gln Asn Thr Tyr Ile Leu Phe Phe Phe
35 40 45

Phe Ser Glu Met Ser
50

<210> 104
<211> 131
<212> PRT
<213> Homo sapien

<400> 104

Met Arg Val Ser Glu Arg Ala Leu Lys Asn Val Ala Cys Gln Gln His
1 5 10 15

Met Asp Ser Leu Phe Arg Val Cys Ile Tyr Pro Ala Asp Thr Pro Ile
20 25 30

Pro Pro Ser Leu Pro Pro Arg Ala Ser Asp Phe Leu Phe His Pro Ala
35 40 45

Ala Tyr Tyr Trp Gln Gly Met Ala Gly Val Asn Leu Gly Ser Val Tyr
50 55 60

His Gln Gly Lys Leu Pro Ser Leu Leu Gln Ser Leu Trp Lys Gly Thr
65 70 75 80

Phe Phe Arg Val Gln His Val Pro Met His Ser Gln Val Pro Lys Val
85 90 95

Thr Tyr Thr Tyr Ile Val Asn Ile Val Pro Thr Ala Leu Gln Thr Phe
100 105 110

Ile Trp Pro Leu Ala Val His Thr Ser Gln Pro Ile His Val Phe Met
115 120 125

Met Met Phe
130

<210> 105
<211> 117
<212> PRT
<213> Homo sapien

<400> 105

Met Ser Ser Phe Gln Gly Phe Ile Phe Gly Gly Lys Lys Ile Pro Gln
1 5 10 15

Asp Ala Gly Cys Pro Ala Ser His Asn Gly Tyr Ala Pro Ile Glu Thr
20 25 30

Ser Ser Gly Arg Val Thr Lys Leu Lys Arg Lys Gln Phe Gln Ala Glu
35 40 45

Gly His Lys Leu Arg Ala Glu Ser Leu Leu Leu Thr Ala Ile Gln Ala
50 55 60

Gln Gly Leu Cys Gly Ala Gly Phe Leu Lys Ala Gly Leu Tyr Leu Gly

65 70 75 80

Arg Arg Glu Arg Thr Arg Gly Leu Asp Ala Gly Trp Arg Phe Cys Asp
85 90 95

Leu Leu Cys Tyr Lys Phe Lys Asn Lys Thr Cys Trp Ile Arg Ser Phe
100 105 110

Ser Tyr Leu Leu Lys
115

<210> 106
<211> 93
<212> PRT
<213> Homo sapien

<400> 106

Met Pro Gly Val Thr Val Lys Asp Val Asn Gln Gln Glu Phe Val Arg
1 5 10 15

Ala Leu Ala Ala Phe Leu Lys Lys Ser Gly Glu Ala Glu Ser Pro Arg
20 25 30

Met Gly Gly Ile Pro Phe Lys Leu Ala Lys Ala Gln Arg Ser Leu Leu
35 40 45

Pro Thr Met Arg Thr Gly Ser Thr Arg Gly Ala Ala Phe Gln Gln Arg
50 55 60

Arg Ala Thr Cys Tyr Leu Pro Gly Val Gly Ala Gly Gly Trp Ala Ser
65 70 75 80

Ile Glu Pro Lys Asp Ser Ile Gly Gly Glu Arg Ser Glu
85 90

<210> 107
<211> 148
<212> PRT
<213> Homo sapien

<400> 107

Met Leu Leu Val Gly Ser Cys His Leu Ser Gly Asp Ser Val Gln Ile
1 5 10 15

Ser Leu Ser Leu Arg Cys Gln Phe Ala Ala Ala Ile Leu Val Leu Phe
20 25 30

His His Phe Gln Pro Leu Gln Gly Leu Glu Asp Pro Ala Gly His Thr
35 40 45

Leu Gly Ala Ser Ala Glu Val Ala Gly His Asp Ala Val Ser Leu Thr
50 55 60

Ser Pro Ile Asp Leu Gly His Gly Ala Asn Pro Ser Ala Thr Pro Glu
65 70 75 80

Val Gln Val Pro Arg Cys Gly Ser Ser Ser Arg Val Glu Pro Val Leu
85 90 95

Ile Val Gly Ser Lys Leu Phe Val Leu Gly Gln Leu Asp Gly Ile His
100 105 110

Pro Phe Gly Asp Phe Gln Leu Pro Gly Leu Phe Glu Glu Gly Cys Gln
115 120 125

Ser Ser Asp Glu Leu Leu Leu Val His Val Phe Tyr Ser Asn Ser Arg
130 135 140

His Arg Ala Ala
145

<210> 108
<211> 172
<212> PRT
<213> Homo sapien

<400> 108

Met Val Cys Gly Gly Phe Ala Cys Ser Ser Leu Arg Val Val Gly Val
1 5 10 15

Val Ile Ala Val Gly Ile Phe Leu Phe Leu Ile Ala Leu Val Gly Leu
20 25 30

Ile Gly Ala Val Lys His His Gln Val Leu Leu Phe Phe Tyr Met Ile
35 40 45

Ile Leu Leu Leu Val Phe Ile Val Gln Phe Ser Val Ser Cys Ala Cys
50 55 60

Leu Ala Leu Asn Gln Glu Gln Gln Gly Gln Leu Leu Glu Val Gly Trp
65 70 75 80

Asn Asn Thr Ala Ser Ala Arg Asn Asp Ile Gln Arg Asn Leu Asn Cys
85 90 95

Cys Gly Phe Arg Ser Val Asn Pro Asn Asp Thr Cys Leu Ala Ser Cys
100 105 110

Val Lys Ser Asp His Ser Cys Ser Pro Cys Ala Pro Ile Ile Gly Glu
115 120 125

Tyr Ala Gly Glu Val Leu Arg Phe Val Gly Gly Ile Gly Leu Phe Phe
130 135 140

Ser Phe Thr Glu Ile Leu Gly Val Trp Leu Thr Tyr Arg Tyr Arg Asn
145 150 155 160

Gln Lys Asp Pro Arg Ala Asn Pro Ser Ala Phe Leu
165 170

<210> 109
<211> 55
<212> PRT
<213> Homo sapien

<400> 109

Met Asn Phe Asp Tyr Ser Val Asn Tyr Trp Asn Val Ser Ser Phe Asn
1 5 10 15

Phe Lys Asn Asn Tyr Phe Thr Ser Ser Asp Trp Gly Phe Pro Glu Ile
20 25 30

Cys Glu Glu Gln Arg Arg Pro Pro Ala Thr Gln His His His Asp Gly
35 40 45

Ala Leu Thr Gly Ser Glu Ser
50 55

<210> 110
<211> 125
<212> PRT
<213> Homo sapien

<400> 110

Met Gln Ala Leu Pro Gln Val Glu Lys Arg Arg Leu Arg Leu Pro Arg
1 5 10 15

Glu Val Gln Cys Pro Ala Leu Leu Arg Arg Met Leu Leu Ile Pro Leu
20 25 30

Trp Lys Ile Pro Ala Pro Thr Thr Thr Lys Ser Cys Arg Glu Thr Phe

35 40 45

Leu Lys Trp Leu Ser Val Ser Ala Ala Glu Arg Thr Thr Gly Ser Trp
50 55 60

Thr Ser Phe Gln Pro Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro
65 70 75 80

Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg
85 90 95

Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser
100 105 110

Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
115 120 125

<210> 111
<211> 1256
<212> PRT
<213> Homo sapien

<400> 111

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser His
50 55 60

Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr Leu
65 70 75 80

Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln
85 90 95

Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr Thr
100 105 110

Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Lys Pro Ala Pro
115 120 125

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
130 135 140

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
145 150 155 160

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
165 170 175

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Pro Pro Gly Ser Thr Ala
180 185 190

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
195 200 205

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
210 215 220

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
225 230 235 240

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
245 250 255

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
260 265 270

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
275 280 285

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
290 295 300

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
305 310 315 320

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
325 330 335

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
340 345 350

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
355 360 365

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
370 375 380

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
385 390 395 400

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
405 410 415

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
420 425 430

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
435 440 445

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
450 455 460

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
465 470 475 480

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
485 490 495

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
500 505 510

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
515 520 525

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
530 535 540

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
545 550 555 560

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
565 570 575

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
580 585 590

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
595 600 605

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr

610 615 620

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
625 630 635 640

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
645 650 655

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
660 665 670

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
675 680 685

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
690 695 700

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
705 710 715 720

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
725 730 735

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
740 745 750

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
755 760 765

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
770 775 780

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
785 790 795 800

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
805 810 815

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala
820 825 830

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
835 840 845

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
850 855 860

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
865 870 875 880

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Val His
885 890 895

Gly Val Thr Ser Ala Pro Asp Ser Arg Ser Gly Ser Gly Phe Leu Pro
900 905 910

Pro Pro Ala Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala
915 920 925

Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp
930 935 940

Asn Arg Pro Ala Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr
945 950 955 960

Ser Ala Ser Gly Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn
965 970 975

Gly Thr Ser Ala Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro
980 985 990

Phe Ser Ile Pro Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser
995 1000 1005

His Ser Thr Lys Thr Asp Ala Ser Ser Thr His His Ser Thr Val
1010 1015 1020

Pro Pro Leu Thr Ser Ser Asn His Ser Thr Ser Pro Gln Leu Ser
1025 1030 1035

Thr Gly Val Ser Phe Phe Phe Leu Ser Phe His Ile Ser Asn Leu
1040 1045 1050

Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp Tyr Tyr Gln
1055 1060 1065

Glu Leu Gln Arg Asp Ile Ser Glu Met Phe Leu Gln Ile Tyr Lys
1070 1075 1080

Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe Arg Pro Gly
1085 1090 1095

Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly Thr Ile
1100 1105 1110

Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr Glu
1115 1120 1125

Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser
1130 1135 1140

Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val Pro
1145 1150 1155

Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala
1160 1165 1170

Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg
1175 1180 1185

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr
1190 1195 1200

Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg
1205 1210 1215

Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val
1220 1225 1230

Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala
1235 1240 1245

Val Ala Ala Thr Ser Ala Asn Leu
1250 1255

<210> 112
<211> 728
<212> PRT
<213> Homo sapien

<400> 112

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser

35 40 45

Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser His
50 55 60

Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr Leu
65 70 75 80

Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln
85 90 95

Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr Thr
100 105 110

Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Lys Pro Ala Pro
115 120 125

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
130 135 140

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Pro Ala His Gly Val Thr
145 150 155 160

Ser Ala Pro Asp Thr Arg Pro Pro Pro Gly Ser Thr Ala Pro Pro Ala
165 170 175

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
180 185 190

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
195 200 205

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
210 215 220

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
225 230 235 240

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
245 250 255

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
260 265 270

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
275 280 285

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
290 295 300

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
305 310 315 320

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
325 330 335

Met Val Ser Ile Gly Leu Ser Phe Pro Ser Ser Pro Glu Ala Ala Ile
340 345 350

Arg Thr Val His Thr Leu Cys Ile Lys Pro Glu Ser Phe Pro Ser His
355 360 365

Pro Ser Phe Cys Arg Phe Ile Asn Lys Gly Val Phe Trp Ala Ser Pro
370 375 380

Ile Leu Ser Ser Gly Thr Val Leu Gly Val Asp Pro Val Trp Trp Leu
385 390 395 400

Glu Gly Trp Val Val Val Met Thr Val Gly Gly Thr Gly Arg Thr Tyr
405 410 415

Gly Trp Gly Lys Ser Arg Glu Pro Glu Leu Gly Pro Val Ala Glu Val
420 425 430

Pro Ile Phe Pro Val Thr Arg Pro Gly Ser Val Val Val Gln Leu Thr
435 440 445

Leu Ala Phe Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln
450 455 460

Phe Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile
465 470 475 480

Ser Asp Val Ser Gly Glu Ala Thr Ser Leu Ala Ala Ala Gln His His
485 490 495

Ala Gly Ala Leu Ser Phe Gln Cys Leu Gly Pro Arg Ser Phe Leu Ser
500 505 510

Ala Gly Ser Gly Arg Gly Ala Ser Ser Gly Arg Leu Pro Cys Pro Leu
515 520 525

Leu Phe Leu Leu Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser
530 535 540

Gly Ala Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys
545 550 555 560

Val Leu Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Val Ser Ala
565 570 575

Val Pro Gly Pro Asp Gln Ser Pro Pro Val Glu Gly Ser Ser Met Ala
580 585 590

Cys His Asn Leu Leu Ser Pro Gln Ala Val Cys Gln Cys Arg Arg Lys
595 600 605

Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro
610 615 620

Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro
625 630 635 640

Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Arg Leu Ala Pro Gln
645 650 655

Ala Arg Gly Ser Arg Gly Phe Gly Trp Ala Arg Ile Leu Lys Gly Val
660 665 670

Leu Gly Lys Pro Lys Glu Leu Gly Arg Gly Glu Lys Trp Arg Glu Val
675 680 685

Ser Arg Gly Gly Pro Gly Lys Asp Glu Gly Gln Arg Ser Glu Glu Phe
690 695 700

Trp Gly Thr Gly Leu Gly Gly Asp Tyr Gly Arg Lys Gly Pro Ser Lys
705 710 715 720

Gly Ser Gly Pro Thr Ala Arg Ile
725

<210> 113
<211> 524
<212> PRT
<213> Homo sapien

<400> 113

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr

1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Thr Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala
165 170 175

Pro Gly Ser Thr Ala Pro Ala Ala His Gly Val Thr Ser Ala Pro Asp
180 185 190

Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr
195 200 205

Ser Ala Pro Asp Asn Arg Pro Ala Leu Gly Ser Thr Ala Pro Pro Val
210 215 220

His Asn Val Thr Ser Ala Ser Gly Ser Ala Ser Gly Ser Ala Ser Thr
225 230 235 240

Leu Val His Asn Gly Thr Ser Ala Arg Ala Thr Thr Thr Pro Ala Ser
245 250 255

Lys Ser Thr Pro Phe Ser Ile Pro Ser His His Ser Asp Thr Pro Thr
260 265 270

Thr Leu Ala Ser His Ser Thr Lys Thr Asp Ala Ser Ser Thr His His
275 280 285

Ser Thr Val Pro Pro Leu Thr Ser Ser Asn His Ser Thr Ser Pro Gln
290 295 300

Leu Ser Thr Gly Val Ser Phe Phe Phe Leu Ser Phe His Ile Ser Asn
305 310 315 320

Leu Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp Tyr Tyr Gln
325 330 335

Glu Leu Gln Arg Asp Ile Ser Glu Met Phe Leu Gln Ile Tyr Lys Gln
340 345 350

Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe Arg Pro Gly Ser Val
355 360 365

Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly Thr Ile Asn Val His
370 375 380

Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg
385 390 395 400

Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser Asp Val Pro Phe Pro
405 410 415

Phe Ser Ala Gln Ser Gly Ala Gly Val Pro Gly Trp Gly Ile Ala Leu
420 425 430

Leu Val Leu Val Cys Val Leu Val Ala Leu Ala Ile Val Tyr Leu Ile
435 440 445

Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp
450 455 460

Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr
465 470 475 480

Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser
485 490 495

Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr
500 505 510

Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
515 520

<210> 114
<211> 515
<212> PRT
<213> Homo sapien

<400> 114

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser His
50 55 60

Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr Leu
65 70 75 80

Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln
85 90 95

Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr Thr
100 105 110

Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Lys Pro Ala Pro
115 120 125

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
130 135 140

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
145 150 155 160

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Ala Ala His
165 170 175

Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala

180 185 190

Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala Leu
195 200 205

Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly Ser
210 215 220

Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala Arg
225 230 235 240

Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro Ser
245 250 255

His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys Thr
260 265 270

Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser Ser
275 280 285

Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe Phe
290 295 300

Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu Asp
305 310 315 320

Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met
325 330 335

Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile
340 345 350

Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg
355 360 365

Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr
370 375 380

Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser
385 390 395 400

Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val
405 410 415

Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala
420 425 430

Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg
435 440 445

Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His
450 455 460

Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro
465 470 475 480

Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn
485 490 495

Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser
500 505 510

Ala Asn Leu
515

<210> 115
<211> 109
<212> PRT
<213> Homo sapien

<400> 115

Met Leu Glu Arg Arg Pro Pro Ala Val Arg Arg Pro Gly Leu Thr Ala
1 5 10 15

Pro Ala Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro
20 25 30

Gly Ser Thr Ala Pro Ala Ala His Gly Val Thr Ser Ala Pro Asp Thr
35 40 45

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Phe
50 55 60

Val Pro Arg Thr Ser Gly Arg Arg Leu Ala Leu Phe Leu Val Tyr Val
65 70 75 80

Phe Arg Val Glu Asp Val Val Gln Thr Arg Leu Asp Thr Leu Arg Ile
85 90 95

Ala Lys Tyr Ile Asp Gly Ser Tyr Ala Val Ser Val Cys
100 105

<210> 116
<211> 174
<212> PRT
<213> Homo sapien

<220>
<221> MISC_FEATURE
<222> (167)..(167)
<223> X= any amino acid

<400> 116

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Thr Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Arg Pro
130 135 140

Ser Cys Gly Ser Gly Leu Gly Thr Ala Cys Val Pro Gly Leu Gln Leu
145 150 155 160

Leu Leu Val Gly Ala His Xaa Thr Gln Leu Leu Thr Tyr Asp
165 170

<210> 117
<211> 475
<212> PRT
<213> Homo sapien

<400> 117

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1   5   10   15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
            20   25   30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
        35   40   45

Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser His
    50   55   60

Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr Leu
65   70   75   80

Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln
                85   90   95

Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr Thr
            100   105   110

Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Lys Pro Ala Pro
        115   120   125

Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr
    130   135   140

Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
145   150   155   160

Ala Pro Asp Asn Arg Pro Ala Leu Gly Ser Thr Ala Pro Pro Val His
                165   170   175

Asn Val Thr Ser Ala Ser Gly Ser Ala Ser Gly Ser Ala Ser Thr Leu
            180   185   190

Val His Asn Gly Thr Ser Ala Arg Ala Thr Thr Thr Pro Ala Ser Lys
        195   200   205

Ser Thr Pro Phe Ser Ile Pro Ser His His Ser Asp Thr Pro Thr Thr
    210   215   220

Leu Ala Ser His Ser Thr Lys Thr Asp Ala Ser Ser Thr His His Ser
225   230   235   240

Thr Val Pro Pro Leu Thr Ser Ser Asn His Ser Thr Ser Pro Gln Leu
245 250 255

Ser Thr Gly Val Ser Phe Phe Phe Leu Ser Phe His Ile Ser Asn Leu
260 265 270

Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp Tyr Tyr Gln Glu
275 280 285

Leu Gln Arg Asp Ile Ser Glu Met Phe Leu Gln Ile Tyr Lys Gln Gly
290 295 300

Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe Arg Pro Gly Ser Val Val
305 310 315 320

Val Gln Leu Thr Leu Ala Phe Arg Glu Gly Thr Ile Asn Val His Asp
325 330 335

Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr
340 345 350

Asn Leu Thr Ile Ser Asp Val Ser Val Ser Asp Val Pro Phe Pro Phe
355 360 365

Ser Ala Gln Ser Gly Ala Gly Val Pro Gly Trp Gly Ile Ala Leu Leu
370 375 380

Val Leu Val Cys Val Leu Val Ala Leu Ala Ile Val Tyr Leu Ile Ala
385 390 395 400

Leu Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile
405 410 415

Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr
420 425 430

His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro
435 440 445

Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr
450 455 460

Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
465 470 475

<210> 118
<211> 231
<212> PRT
<213> Homo sapien

<400> 118

Met Cys Pro Leu Ala Val Pro Ile Val Ala Pro Met Arg Arg Phe Leu
1 5 10 15

Gln Val Met Val Ala Ala Ala Ser Leu Thr Gln Thr Gln Gln Trp Gln
20 25 30

Pro Leu Leu Pro Thr Cys Arg Gly Thr Ser Pro Ala Glu Leu Ser Gly
35 40 45

Gln Pro Val Pro Phe His Ser Thr Gln Val Leu Gln Gly Gln Ser Pro
50 55 60

Cys Thr Leu Phe Gly Leu Val Ser Trp Glu Phe Arg Trp Ala Ala His
65 70 75 80

Ser Leu Leu Gln Arg Pro His Asp Tyr Phe Arg Lys Phe Glu Pro His
85 90 95

Leu Tyr Ser Leu Asp Ser Asn Ser Asp Asp Val Asp Ser Leu Thr Asp
100 105 110

Glu Glu Ile Leu Ser Lys Tyr Gln Leu Gly Met Leu His Phe Ser Thr
115 120 125

Gln Tyr Asp Leu Leu His Asn His Leu Thr Val Arg Val Ile Glu Ala
130 135 140

Arg Asp Leu Pro Pro Pro Ile Ser His Asp Gly Ser Arg Gln Asp Met
145 150 155 160

Ala His Ser Asn Pro Tyr Val Lys Ile Cys Leu Leu Pro Asp Gln Lys
165 170 175

Asn Ser Lys Gln Thr Gly Val Lys Arg Lys Thr Gln Lys Pro Val Phe
180 185 190

Glu Glu Arg Tyr Thr Phe Glu Ile Pro Phe Leu Glu Ala Gln Arg Arg
195 200 205

Thr Leu Leu Leu Thr Val Val Asp Phe Asp Lys Phe Ser Arg His Cys

210 215 220

Val Ile Gly Lys Val Ser Val
225 230

<210> 119
<211> 107
<212> PRT
<213> Homo sapien

<400> 119

Met Val Ala Ala Ala Ser Leu Thr Gln Thr Gln Gln Trp Gln Pro Leu
1 5 10 15

Leu Pro Thr Cys Arg Gly Thr Ser Pro Ala Glu Leu Ser Gly Gln Pro
20 25 30

Val Pro Phe His Ser Thr Gln Val Leu Gln Gly Gln Ser Pro Cys Thr
35 40 45

Leu Phe Gly Leu Val Ser Trp Glu Phe Arg Trp Ala Ala His Ser Leu
50 55 60

Leu Gln Arg Pro His Gln Phe Leu Gly His Phe Ser Val Cys Gly Ser
65 70 75 80

Ser Cys Gly Pro Leu Arg Ala His Ala Trp Glu Val Leu Trp Trp Gly
85 90 95

Leu Pro Gly Gly Leu Ala Gln Arg Ala Leu Arg
100 105

<210> 120
<211> 484
<212> PRT
<213> Homo sapien

<400> 120

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu

290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly
370 375 380

Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val
385 390 395 400

Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg
405 410 415

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr
420 425 430

His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val
435 440 445

Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly
450 455 460

Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr
465 470 475 480

Ser Ala Asn Leu

<210> 121
<211> 463
<212> PRT
<213> Homo sapien

<400> 121

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

```
Val Leu Thr Gly Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser
            20                  25                  30

Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val
        35                  40                  45

Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln
    50                  55                  60

Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala
65                  70                  75                  80

Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu
                85                  90                  95

Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn
            100                 105                 110

Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser
        115                 120                 125

Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His
    130                 135                 140

Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala Leu Gly Ser Thr Ala
145                 150                 155                 160

Pro Pro Val His Asn Val Thr Ser Ala Ser Gly Ser Ala Ser Gly Ser
                165                 170                 175

Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala Arg Ala Thr Thr Thr
            180                 185                 190

Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro Ser His His Ser Asp
        195                 200                 205

Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys Thr Asp Ala Ser Ser
    210                 215                 220

Thr His His Ser Thr Val Pro Pro Leu Thr Ser Ser Asn His Ser Thr
225                 230                 235                 240

Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe Phe Leu Ser Phe His
                245                 250                 255
```

Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp
260 265 270

Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met Phe Leu Gln Ile
275 280 285

Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe Arg Pro
290 295 300

Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly Thr Ile
305 310 315 320

Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr Glu Ala
325 330 335

Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser Asp Val
340 345 350

Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val Pro Gly Trp Gly
355 360 365

Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala Leu Ala Ile Val
370 375 380

Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly
385 390 395 400

Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu
405 410 415

Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr
420 425 430

Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser
435 440 445

Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
450 455 460

<210> 122
<211> 524
<212> PRT
<213> Homo sapien

<400> 122

Met Gly Arg Glu Lys Glu Ala Ala Ala Gly Lys Glu Ala Ala Asn Pro
1 5 10 15

Gly Val Thr Glu Ala Ala His Ser Pro Val Leu Leu Val Leu Phe Leu
20 25 30

Trp Trp Pro Glu Leu Ile Phe Ser Ser Cys Ser Tyr Phe Ser Phe Ile
35 40 45

Lys Thr Gln Pro Tyr Asp Phe Asn Phe Phe Thr Ala Thr Thr Ala Pro
50 55 60

Lys Pro Ala Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro
65 70 75 80

Gly Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser
85 90 95

Ser Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser
100 105 110

His Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr
115 120 125

Leu Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly
130 135 140

Gln Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr
145 150 155 160

Thr Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Lys Pro Ala
165 170 175

Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp
180 185 190

Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr
195 200 205

Ser Ala Pro Asp Asn Arg Pro Ala Leu Gly Ser Thr Ala Pro Pro Val
210 215 220

His Asn Val Thr Ser Ala Ser Gly Ser Ala Ser Gly Ser Ala Ser Thr
225 230 235 240

Leu Val His Asn Gly Thr Ser Ala Arg Ala Thr Thr Thr Pro Ala Ser
245 250 255

Lys Ser Thr Pro Phe Ser Ile Pro Ser His His Ser Asp Thr Pro Thr
260 265 270

Thr Leu Ala Ser His Ser Thr Lys Thr Asp Ala Ser Ser Thr His His
275 280 285

Ser Thr Val Pro Pro Leu Thr Ser Ser Asn His Ser Thr Ser Pro Gln
290 295 300

Leu Ser Thr Gly Val Ser Phe Phe Phe Leu Ser Phe His Ile Ser Asn
305 310 315 320

Leu Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp Tyr Tyr Gln
325 330 335

Glu Leu Gln Arg Asp Ile Ser Glu Met Phe Leu Gln Ile Tyr Lys Gln
340 345 350

Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe Arg Pro Gly Ser Val
355 360 365

Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly Thr Ile Asn Val His
370 375 380

Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg
385 390 395 400

Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser Asp Val Pro Phe Pro
405 410 415

Phe Ser Ala Gln Ser Gly Ala Gly Val Pro Gly Trp Gly Ile Ala Leu
420 425 430

Leu Val Leu Val Cys Val Leu Val Ala Leu Ala Ile Val Tyr Leu Ile
435 440 445

Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp
450 455 460

Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr
465 470 475 480

Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser
485 490 495

Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr
500 505 510

Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
515 520

<210> 123
<211> 435
<212> PRT
<213> Homo sapien

<400> 123

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser His
50 55 60

Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr Leu
65 70 75 80

Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln
85 90 95

Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr Thr
100 105 110

Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Arg Pro Ala Leu
115 120 125

Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly Ser
130 135 140

Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala Arg
145 150 155 160

Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro Ser
165 170 175

His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys Thr
180 185 190

Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser Ser
195 200 205

Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe Phe
210 215 220

Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu Asp
225 230 235 240

Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met
245 250 255

Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile
260 265 270

Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg
275 280 285

Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr
290 295 300

Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser
305 310 315 320

Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val
325 330 335

Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala
340 345 350

Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg
355 360 365

Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His
370 375 380

Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro
385 390 395 400

Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn
405 410 415

Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser
420 425 430

Ala Asn Leu
435

<210> 124
<211> 273
<212> PRT
<213> Homo sapien

<400> 124

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Leu Ser Thr Gly Val Ser Phe Phe Phe Leu Ser
50 55 60

Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser
65 70 75 80

Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met Phe Leu
85 90 95

Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe
100 105 110

Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly
115 120 125

Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr
130 135 140

Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser
145 150 155 160

Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val Pro Gly
165 170 175

Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala Leu Ala
180 185 190

Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn

195 200 205

Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met
210 215 220

Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser
225 230 235 240

Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly
245 250 255

Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn
260 265 270

Leu

<210> 125
<211> 350
<212> PRT
<213> Homo sapien

<400> 125

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
    130   135   140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145   150   155   160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
                165   170   175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
            180   185   190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
        195   200   205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
    210   215   220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225   230   235   240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
                245   250   255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
            260   265   270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
        275   280   285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
    290   295   300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305   310   315   320

Ile Lys Phe Arg Tyr Ser Ser Gly Cys Gly Pro Ser Val Val Val Gly
                325   330   335

Gly Trp Val Val Val Met Thr Val Gly Arg Asp Trp Cys Thr
            340   345   350

<210> 126
<211> 316
<212> PRT
<213> Homo sapien

<400> 126

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Val Ser Ile Gly Leu Ser Phe Pro Met Leu Pro
305 310 315

<210> 127
<211> 230
<212> PRT
<213> Homo sapien

<400> 127

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Ile Pro Ala Pro Thr Thr Thr Lys Ser Cys Arg
50 55 60

Glu Thr Phe Leu Lys Trp Pro Gly Ser Val Val Val Gln Leu Thr Leu
65 70 75 80

Ala Phe Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe
85 90 95

Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser
100 105 110

Asp Val Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly
115 120 125

Ala Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val
130 135 140

Leu Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln
145 150 155 160

Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp
165 170 175

Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg
180 185 190

Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser
195 200 205

Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala
210 215 220

Ala Thr Ser Ala Asn Leu
225 230

<210> 128
<211> 614
<212> PRT
<213> Homo sapien

<400> 128

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Val Leu Leu Ile Gly Gly Gly Glu Arg Arg Tyr Arg Ala Met Val
370 375 380

Ser Ala Thr Gly Ile Ser Leu Gly Ala Met Ala Gly Lys Gly Gly Gly
385 390 395 400

Val Ser Glu Trp Trp Leu Gly Ile Glu Asn Gly Val Leu Leu Leu Ala
405 410 415

Gly Val Val Val Ala Leu Ala Glu Val Pro Leu Cys Thr Arg Val Glu
420 425 430

Ala Glu Pro Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu
435 440 445

Thr Ser Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser
450 455 460

Phe Phe Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser
465 470 475 480

Leu Glu Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile
485 490 495

Ser Glu Asp Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly
500 505 510

Ala Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val
515 520 525

Leu Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln
530 535 540

Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp
545 550 555 560

Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg
565 570 575

Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser
580 585 590

Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala
595 600 605

Ala Thr Ser Ala Asn Leu
610

<210> 129
<211> 372
<212> PRT
<213> Homo sapien

<400> 129

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Trp Gly Ala Arg Leu Gly His Arg Ala Ala Gly Ala Gly Leu Cys
305 310 315 320

Ser Gly Cys Ala Gly His Cys Leu Ser His Cys Leu Gly Cys Leu Ser
325 330 335

Val Pro Pro Lys Glu Leu Arg Ala Ala Gly His Leu Ser Ser Pro Gly
340 345 350

Tyr Leu Pro Ser Tyr Glu Arg Val Pro His Leu Pro His Pro Trp Ala
355 360 365

Leu Cys Ala Pro
370

<210> 130
<211> 256
<212> PRT
<213> Homo sapien

<400> 130

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Val Ser Met Thr Ser Ser Val Leu Ser Ser His
50 55 60

Ser Pro Gly Ser Gly Ser Ser Thr Thr Gln Gly Gln Asp Val Thr Leu
65 70 75 80

Ala Pro Ala Thr Glu Pro Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln
85 90 95

Asp Val Thr Ser Val Pro Val Thr Arg Pro Ala Leu Gly Ser Thr Thr
100 105 110

Pro Pro Ala His Asp Val Thr Ser Ala Pro Asp Asn Arg Pro Ala Leu
115 120 125

Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly Ser
130 135 140

Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala Arg
145 150 155 160

Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro Ser
165 170 175

His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys Thr
180 185 190

Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser Ser
195 200 205

Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe Phe
210 215 220

Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu Asp
225 230 235 240

Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met
245 250 255

<210> 131

<211> 492
<212> PRT
<213> Homo sapien

<400> 131

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly
370 375 380

Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val
385 390 395 400

Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg
405 410 415

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr
420 425 430

His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val
435 440 445

Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Arg Leu Gly
450 455 460

Pro Thr Gly Gln Gly Lys Gln Arg Val Trp Leu Gly Lys Asp Ser Glu
465 470 475 480

Gly Gly Thr Trp Lys Thr Gln Arg Ala Trp Lys Arg
485 490

<210> 132
<211> 483
<212> PRT
<213> Homo sapien

<400> 132

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly
370 375 380

Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val
385 390 395 400

Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg
405 410 415

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr
420 425 430

His Pro Met Ser Glu Trp Arg Val Tyr Glu Glu Lys Lys Lys Glu Val
435 440 445

Pro Ala Val Pro Glu Thr Leu Lys Lys Lys Arg Arg Asn Phe Ala Glu
450 455 460

Leu Lys Ile Lys Arg Leu Arg Lys Lys Phe Ala Lys Arg Cys Phe Glu
465 470 475 480

Arg Gln Gly

<210> 133
<211> 150
<212> PRT
<213> Homo sapien

<400> 133

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp
50 55 60

Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met Ala Val Cys Gln
65 70 75 80

Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp
85 90 95

Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg
100 105 110

Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser
115 120 125

Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala
130 135 140

Ala Thr Ser Ala Asn Leu

145 150

<210> 134
<211> 168
<212> PRT
<213> Homo sapien

<400> 134

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Leu Ser Thr Gly Val Ser Phe Phe Phe Leu Ser
50 55 60

Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu Asp Pro Ser
65 70 75 80

Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met Ala Val
85 90 95

Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala
100 105 110

Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His
115 120 125

Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys
130 135 140

Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala
145 150 155 160

Val Ala Ala Thr Ser Ala Asn Leu
165

<210> 135
<211> 79
<212> PRT
<213> Homo sapien

<400> 135

Ser Pro Glu Trp Leu Thr Leu Ile Ser Ser Pro Gly Lys Asn Tyr Gly
1 5 10 15

Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu
20 25 30

Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr
35 40 45

Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser
50 55 60

Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
65 70 75

<210> 136
<211> 398
<212> PRT
<213> Homo sapien

<400> 136

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Ala Glu Val Pro Phe His Ile Met Leu Thr Asn Met Gly Thr Met
370 375 380

Glu Tyr His Asn Val Gly Ala Ile Arg Phe Arg His Asn Tyr
385 390 395

<210> 137
<211> 36
<212> PRT
<213> Homo sapien

<400> 137

Gly Arg Leu Leu Leu Leu Leu Leu Glu Phe Lys Leu Leu Thr Met Tyr
1 5 10 15

Gly Leu Met Pro Gly Lys Cys Cys Gly Gly Gly Ser Gln Glu Asp Trp
20 25 30

Pro Arg Glu Pro
35

<210> 138
<211> 264
<212> PRT
<213> Homo sapien

<400> 138

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Phe Asn
50 55 60

Ser Ser Leu Glu Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg
65 70 75 80

Asp Ile Ser Glu Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu
85 90 95

Gly Leu Ser Asn Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu
100 105 110

Thr Leu Ala Phe Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr
115 120 125

```
Gln Phe Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr
    130                 135                 140

Ile Ser Asp Val Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln
145                 150                 155                 160

Ser Gly Ala Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val
                165                 170                 175

Cys Val Leu Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val
            180                 185                 190

Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala
        195                 200                 205

Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His
    210                 215                 220

Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys
225                 230                 235                 240

Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala
                245                 250                 255

Val Ala Ala Thr Ser Ala Asn Leu
            260


<210>  139
<211>  241
<212>  PRT
<213>  Homo sapien

<400>  139

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1               5                   10                  15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
            20                  25                  30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
        35                  40                  45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Phe Leu
    50                  55                  60
```

Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe
65 70 75 80

Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly
85 90 95

Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr
100 105 110

Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser
115 120 125

Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val Pro Gly
130 135 140

Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala Leu Ala
145 150 155 160

Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn
165 170 175

Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met
180 185 190

Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser
195 200 205

Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly
210 215 220

Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn
225 230 235 240

Leu

<210> 140
<211> 92
<212> PRT
<213> Homo sapien

<400> 140

Met Ala Cys His Asn Leu Leu Ser Pro Gln Ala Val Cys Gln Cys Arg
1 5 10 15

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr
20 25 30

His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val
35 40 45

Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Arg Leu Gly
50 55 60

Pro Thr Gly Gln Gly Lys Gln Arg Val Trp Leu Gly Lys Asp Ser Glu
65 70 75 80

Gly Gly Thr Trp Lys Thr Gln Arg Ala Trp Lys Arg
85 90

<210> 141
<211> 420
<212> PRT
<213> Homo sapien

<400> 141

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
1 5 10 15

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
20 25 30

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
35 40 45

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
50 55 60

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
65 70 75 80

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
85 90 95

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
100 105 110

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
115 120 125

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
130 135 140

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
145 150 155 160

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
165 170 175

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
180 185 190

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
195 200 205

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
210 215 220

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
225 230 235 240

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
245 250 255

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
260 265 270

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
275 280 285

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
290 295 300

Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly
305 310 315 320

Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val
325 330 335

Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg
340 345 350

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr
355 360 365

His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val
370 375 380

Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly

385 390 395 400

Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr
405 410 415

Ser Ala Asn Leu
420

<210> 142
<211> 485
<212> PRT
<213> Homo sapien

<400> 142

Met Pro Gln Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu
1 5 10 15

Thr Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr
20 25 30

Gly Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser
35 40 45

Ala Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val
50 55 60

Ser Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser
65 70 75 80

Ser Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro
85 90 95

Ala Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro
100 105 110

Val Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val
115 120 125

Thr Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro
130 135 140

Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser
145 150 155 160

Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro
165 170 175

Ala Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser
180 185 190

Gly Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser
195 200 205

Ala Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile
210 215 220

Pro Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr
225 230 235 240

Lys Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr
245 250 255

Ser Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe
260 265 270

Phe Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu
275 280 285

Glu Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser
290 295 300

Glu Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser
305 310 315 320

Asn Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala
325 330 335

Phe Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn
340 345 350

Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp
355 360 365

Val Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala
370 375 380

Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu
385 390 395 400

Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys
405 410 415

Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr
420 425 430

Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr
435 440 445

Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala
450 455 460

Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala
465 470 475 480

Thr Ser Ala Asn Leu
485

<210> 143
<211> 255
<212> PRT
<213> Homo sapien

<400> 143

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Phe Asn Ser Ser Leu Glu Asp Pro Ser Thr Asp
50 55 60

Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu Met Phe Leu Gln Ile
65 70 75 80

Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn Ile Lys Phe Arg Pro
85 90 95

Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe Arg Glu Gly Thr Ile
100 105 110

Asn Val His Asp Val Glu Thr Gln Phe Asn Gln Tyr Lys Thr Glu Ala
115 120 125

Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val Ser Val Ser Asp Val
130 135 140

Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly Val Pro Gly Trp Gly
145 150 155 160

Ile Ala Leu Leu Val Leu Val Cys Val Leu Val Ala Leu Ala Ile Val
165 170 175

Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly
180 185 190

Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu
195 200 205

Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr
210 215 220

Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser
225 230 235 240

Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
245 250 255

<210> 144
<211> 517
<212> PRT
<213> Homo sapien

<400> 144

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Val Ser Ile Gly Leu Ser Phe Pro Ser Ser Pro Glu Ala Ala Ile
305 310 315 320

Arg Thr Val His Thr Leu Cys Ile Lys Pro Glu Ser Phe Pro Ser His
325 330 335

Pro Ser Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser

340 345 350

Asn Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala
355 360 365

Phe Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn
370 375 380

Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp
385 390 395 400

Val Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala
405 410 415

Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu
420 425 430

Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys
435 440 445

Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr
450 455 460

Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr
465 470 475 480

Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala
485 490 495

Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala
500 505 510

Thr Ser Ala Asn Leu
515

<210> 145
<211> 180
<212> PRT
<213> Homo sapien

<400> 145

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
1 5 10 15

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
20 25 30

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
35 40 45

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
50 55 60

Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln Ser Gly Ala Gly
65 70 75 80

Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val Cys Val Leu Val
85 90 95

Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val Cys Gln Cys Arg
100 105 110

Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr
115 120 125

His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val
130 135 140

Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly
145 150 155 160

Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr
165 170 175

Ser Ala Asn Leu
180

<210> 146
<211> 232
<212> PRT
<213> Homo sapien

<400> 146

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu

50 55 60

Gly Leu Ser Asn Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu
65 70 75 80

Thr Leu Ala Phe Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr
85 90 95

Gln Phe Asn Gln Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr
100 105 110

Ile Ser Asp Val Ser Val Ser Asp Val Pro Phe Pro Phe Ser Ala Gln
115 120 125

Ser Gly Ala Gly Val Pro Gly Trp Gly Ile Ala Leu Leu Val Leu Val
130 135 140

Cys Val Leu Val Ala Leu Ala Ile Val Tyr Leu Ile Ala Leu Ala Val
145 150 155 160

Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile Phe Pro Ala
165 170 175

Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr His Thr His
180 185 190

Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro Tyr Glu Lys
195 200 205

Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr Asn Pro Ala
210 215 220

Val Ala Ala Thr Ser Ala Asn Leu
225 230

<210> 147
<211> 396
<212> PRT
<213> Homo sapien

<400> 147

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Gly Glu Ala Thr Ser Leu Ala Ala Ala Gln His His Ala Gly Ala
370 375 380

Pro Leu Leu Pro Val Ser Gly Ser Pro Leu Phe Pro
385 390 395

<210> 148
<211> 325
<212> PRT
<213> Homo sapien

<400> 148

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Ser Glu
325

<210> 149
<211> 409
<212> PRT
<213> Homo sapien

<400> 149

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1   5   10   15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
            20   25   30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
        35   40   45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
    50   55   60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65   70   75   80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
                85   90   95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
            100   105   110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
        115   120   125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
    130   135   140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145   150   155   160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
                165   170   175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
            180   185   190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
        195   200   205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro

210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305 310 315 320

Ile Lys Phe Arg Pro Gly Ser Val Val Val Gln Leu Thr Leu Ala Phe
325 330 335

Arg Glu Gly Thr Ile Asn Val His Asp Val Glu Thr Gln Phe Asn Gln
340 345 350

Tyr Lys Thr Glu Ala Ala Ser Arg Tyr Asn Leu Thr Ile Ser Asp Val
355 360 365

Ser Gly Cys Leu Ser Val Pro Pro Lys Glu Leu Arg Ala Ala Gly His
370 375 380

Leu Ser Ser Pro Gly Tyr Leu Pro Ser Tyr Glu Arg Val Pro His Leu
385 390 395 400

Pro His Pro Trp Ala Leu Cys Ala Pro
405

<210> 150
<211> 379
<212> PRT
<213> Homo sapien

<400> 150

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala
35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
290 295 300

Met Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln Leu Asp Ile
305 310 315 320

Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu Tyr Pro Thr Tyr
325 330 335

His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp Arg Ser Pro
340 345 350

Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser Leu Ser Tyr Thr
355 360 365

Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
370 375

<210> 151
<211> 110
<212> PRT
<213> Homo sapien

<400> 151

Val Val Thr Trp His Asn Pro Gly Ala Gly Val Pro Gly Trp Gly Ile
1 5 10 15

Ala Leu Leu Val Leu Val Cys Val Leu Val Ala Leu Ala Ile Val Tyr
20 25 30

Leu Ile Ala Leu Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly Gln
35 40 45

Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu Tyr
50 55 60

Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr Asp
65 70 75 80

Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser Leu
85 90 95

Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
100 105 110

<210> 152
<211> 127
<212> PRT
<213> Homo sapien

<400> 152

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Val Val Thr Gly Ser Gly His Ala Ser Ser Thr Pro Gly
20 25 30

Gly Glu Lys Glu Thr Ser Ala Thr Gln Arg Ser Ser Val Pro Ser Ser
35 40 45

Thr Glu Lys Asn Ala Ala Val Cys Gln Cys Arg Arg Lys Asn Tyr Gly
50 55 60

Gln Leu Asp Ile Phe Pro Ala Arg Asp Thr Tyr His Pro Met Ser Glu
65 70 75 80

Tyr Pro Thr Tyr His Thr His Gly Arg Tyr Val Pro Pro Ser Ser Thr
85 90 95

Asp Arg Ser Pro Tyr Glu Lys Val Ser Ala Gly Asn Gly Gly Ser Ser
100 105 110

Leu Ser Tyr Thr Asn Pro Ala Val Ala Ala Thr Ser Ala Asn Leu
115 120 125

<210> 153
<211> 336
<212> PRT
<213> Homo sapien

<400> 153

Met Thr Pro Gly Thr Gln Ser Pro Phe Phe Leu Leu Leu Leu Leu Thr
1 5 10 15

Val Leu Thr Ala Thr Thr Ala Pro Lys Pro Ala Thr Val Val Thr Gly
20 25 30

Ser Gly His Ala Ser Ser Thr Pro Gly Gly Glu Lys Glu Thr Ser Ala

35 40 45

Thr Gln Arg Ser Ser Val Pro Ser Ser Thr Glu Lys Asn Ala Val Ser
50 55 60

Met Thr Ser Ser Val Leu Ser Ser His Ser Pro Gly Ser Gly Ser Ser
65 70 75 80

Thr Thr Gln Gly Gln Asp Val Thr Leu Ala Pro Ala Thr Glu Pro Ala
85 90 95

Ser Gly Ser Ala Ala Thr Trp Gly Gln Asp Val Thr Ser Val Pro Val
100 105 110

Thr Arg Pro Ala Leu Gly Ser Thr Thr Pro Pro Ala His Asp Val Thr
115 120 125

Ser Ala Pro Asp Asn Lys Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala
130 135 140

His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr
145 150 155 160

Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Asn Arg Pro Ala
165 170 175

Leu Gly Ser Thr Ala Pro Pro Val His Asn Val Thr Ser Ala Ser Gly
180 185 190

Ser Ala Ser Gly Ser Ala Ser Thr Leu Val His Asn Gly Thr Ser Ala
195 200 205

Arg Ala Thr Thr Thr Pro Ala Ser Lys Ser Thr Pro Phe Ser Ile Pro
210 215 220

Ser His His Ser Asp Thr Pro Thr Thr Leu Ala Ser His Ser Thr Lys
225 230 235 240

Thr Asp Ala Ser Ser Thr His His Ser Thr Val Pro Pro Leu Thr Ser
245 250 255

Ser Asn His Ser Thr Ser Pro Gln Leu Ser Thr Gly Val Ser Phe Phe
260 265 270

Phe Leu Ser Phe His Ile Ser Asn Leu Gln Phe Asn Ser Ser Leu Glu
275 280 285

```
Asp Pro Ser Thr Asp Tyr Tyr Gln Glu Leu Gln Arg Asp Ile Ser Glu
    290                 295                 300

Met Phe Leu Gln Ile Tyr Lys Gln Gly Gly Phe Leu Gly Leu Ser Asn
305                 310                 315                 320

Ile Lys Phe Ser Gln Glu Leu Trp Trp Gln Asn Lys Arg Ser Ser Asn
                325                 330                 335


<210>  154
<211>  55
<212>  PRT
<213>  Homo sapien

<400>  154

Met Ala Thr Gln Leu Ile Leu Val Gln Met Ser Leu Phe Pro Asp Ala
1               5                   10                  15

Pro His Asp Pro Ser Ser Leu Gly Gly Met His Pro Ser Ser Val Ser
            20                  25                  30

His Phe Arg Ala Phe Cys Thr Leu Leu Thr Leu Ser Arg Ile Pro Ala
        35                  40                  45

Ile Trp Val Gln Ala Ser Gln
    50                  55


<210>  155
<211>  97
<212>  PRT
<213>  Homo sapien

<400>  155

Met Asn His Leu Arg His Phe Cys Ile Thr Glu Asp Leu Ala Leu Pro
1               5                   10                  15

Ala Val Leu Gly Met Leu Arg Val Gln Ala Glu Ala Arg Glu Ala Gly
            20                  25                  30

Leu Arg Gly Gln Lys Gln Gly Leu Gly Gln Ala Ser Pro Asp Arg Glu
        35                  40                  45

Glu Gly Leu Arg Ser Lys Pro Ala Val Leu Leu Ala Gly Gly Pro Gly
    50                  55                  60

Gly Tyr Lys Leu Leu Cys Ala Leu His Lys Pro Glu Ser Pro Thr Arg
```

65 70 75 80

Glu Asp Val Cys Glu Glu Pro Glu Thr Leu Asn Pro Ile Lys Arg Glu
85 90 95

Lys

<210> 156
<211> 52
<212> PRT
<213> Homo sapien

<400> 156

Met Leu Cys Ala Ile Ser Ile Ser Leu Val Ile Phe Phe Asn Lys His
1 5 10 15

Glu Ser Ile Lys Lys Lys Arg Arg Lys Lys Lys Ala Gly Gly Thr Leu
20 25 30

Gly Gln Ser Gly Pro Gly Gly Asp Trp Phe Pro Ala Asn Ser Thr Ile
35 40 45

Ser Arg Thr Arg
50

<210> 157
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 157
cacttccttt agttttgccc tgg 23

<210> 158
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 158
atcctgaatt ctgagaccat cca 23

<210> 159
<211> 21
<212> DNA

<213> Artificial sequence

<220>
<223> Synthetic

<400> 159
gcctccagca cactcttcag t 21

<210> 160
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 160
agccggagga gatgtggctc taccg 25

<210> 161
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 161
ccgcttccca gagactcatc 20

<210> 162
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 162
gcacaaacat cggcttggt 19

<210> 163
<211> 27
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 163
agagagacat ttctgaaatg gctgtct 27

<210> 164
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 164
cccagcaccg actactacca a 21

<210> 165
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 165
agctgcccgt agttctttcg 20

<210> 166
<211> 27
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 166
ctgaaagcag gtcacccctg agatcct 27

<210> 167
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 167
cagagcttgg ccaggttcta a 21

<210> 168
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 168
tgctagggtg cccctctgt 19

<210> 169
<211> 24
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 169
cctttagggc ctgggacaac cacg 24

<210> 170
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 170
tggataacaa gcccacaaat ga 22

<210> 171
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic

<400> 171
cctctagttc cagccccttt tag 23

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 2003/106648 A3

(51) International Patent Classification[7]: C12N 15/12, 15/11, 15/00, C12P 21/02, C12Q 1/68, A61K 48/00

(21) International Application Number: PCT/US2003/018934

(22) International Filing Date: 16 June 2003 (16.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/389,327 14 June 2002 (14.06.2002) US

(71) Applicant (*for all designated States except US*): **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **SALCEDA, Susana** [AR/US]; 4118 Cresendo Avenue, San Jose, CA 95136 (US). **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **TURNER, Leah, R.** [US/US]; 939 Rosette Court, Sunnyvale, CA 94086 (US). **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report: 7 October 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2003/106648 A3

(54) Title: COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions containing the nucleic acid molecules, polypeptides, antibodies, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/18934

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/12, 15/11, 15/00; C12P 21/02; C12Q 1/68; A61K 48/00
US CL : 536/23.5, 23.1; 435/6, 320.1, 325, 252.3, 69.1; 514/44
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.5, 23.1; 435/6, 320.1, 325, 252.3, 69.1; 514/44

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Compugen, SEQ ID NO: 95

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database LOCUS on EST, AN BI825997, 'National Institutes of Health, Mammalian Gene Collection (MGC),' Gene Sequence, 04 October 2001.	1, 2, 4-6, 8, 9, and 16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
01 April 2004 (01.04.2004)

Date of mailing of the international search report
20 AUG 2004

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)872-9306

Authorized officer
James Martinell
Dorothea Lawrence for

Telephone No. (571) 272-1600

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/18934

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-10 and 15-18 as they pertain to nucleic acids that encode SEQ ID NO: 95

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US03/18934

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group I, claim(s) 1-10 and 15-18, drawn to nucleic acids, nucleic acid molecular hybridization assays, vectors, host cells, methods for producing proteins, kits, and methods of treatment using nucleic acids.

Group II, claim(s) 11, 12, 17, and 18, drawn to polypeptides, kits, methods of treatment using polypeptides, and polypeptide vaccines.

Group III, claim(s) 13-16, drawn to antibodies and protein binding assays.

The inventions listed as Groups I-III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The nucleic acids, vectors, host cells, vaccines, and kits of Group I are materially different from the polypeptides, vaccines, and kits of Group II and the antibodies of Group III. The methods of Group I may be practiced without the polypeptides or vaccines of Group II and without the antibodies of Group III. The methods of Groups I-III may be practiced independently of one another. The polypeptides, vaccines, and kits of Group II are materially different from the antibodies of Group III and are not needed to practice the methods of Group III.

Each of the Groups mentions or requires the use of a large number of separate and unrelated nucleic acids and/or polypeptides. No matter which additional Group(s) applicant elects, applicant is further required to select for search one SEQ ID NO within the Group(s) for search. Any additional SEQ ID NO to be searched requires one additional search fee per SEQ ID NO. In the absence of payment of additional search fee(s) the first mentioned SEQ ID NO in Group I will be searched. Should applicant pay fee(s) for additional Groups to be searched, the first mentioned SEQ ID NO within the selected |g will be searched unless applicant directs otherwise.